I hereby certify that this correspondence is being deposited with the United States Patent and Trademark Office via EFS on **September 28, 2009**.

Michelle Hobson

Michelle Hobson

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:	Examiner: T. Wessendorf
WANG and PABO	Group Art Unit: 1639
For: **DIMERIZING PEPTIDES**	Confirmation No.: 6438
Serial No.: 09/636,243	
Filed: August 10, 2000	
Atty. Docket No.: 8325-1004 (M4-US1)	

BRIEF ON APPEAL UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is filed pursuant to 37 C.F.R. § 41.37 and is in response to the Final Office Action mailed May 1, 2009 and an Advisory Action mailed July 13, 2009. A Notice of Appeal was received in the USPTO on July 30, 2009, making an Appeal Brief due on or before September 30, 2009. Accordingly, this Appeal Brief is timely filed.

REAL PARTIES IN INTEREST

The Massachusetts Institute of Technology, the assignee of record of the above-referenced patent application by virtue of an assignment recorded on April 2, 2001 at Reel 011676, Frame 0049. Sangamo BioSciences, Inc., is the exclusive licensee of the above-referenced patent application. Thus, the Massachusetts Institute of Technology and Sangamo BioSciences, Inc. are the real parties in interest in this matter.

RELATED APPEALS AND INTERFERENCES

Appellants note that this case was previously appealed. The Board decision reversing the previous rejections and setting forth new grounds of rejection was mailed on May 30, 2007 and is attached hereto.

STATUS OF THE CLAIMS

Pending: claims 5, 6, 20 and 21

Canceled: claims 1-4 and 7-19

Rejected: claims 5, 6, 20 and 21

Appealed: claims 5, 6, 20 and 21

STATUS OF THE AMENDMENTS

The Advisory Action indicated amendments made after final would not be entered. However, Appellants note that no amendments were made to the claims in the response to the Final Office Action mailed May 1, 2009. Thus, the claims on appeal are as shown in the Claims Appendix.

SUMMARY OF THE CLAIMED SUBJECT MATTER

Independent claim 5 is drawn to a zinc finger complex, comprising two or more fusion proteins (page 3, lines 26-29). Each fusion protein comprising a zinc finger protein that binds to DNA in a sequence-specific manner (page 1, lines 11-12) and a non-naturally occurring peptide linker of 30 amino acids or less in length (page 3, lines 24-25; page 3, lines 29-30; page 4, lines 14-15). The zinc finger proteins of the fusion proteins are joined to each other by specific binding of the peptide linkers (page 3, lines 29-30).

Claim 6 depends from claim 5 and further indicates that the peptide linker of each fusion protein is the same (page 3, lines 21-32).

Claim 20 depends from claim 5 and further specifies that the zinc finger protein of each fusion protein has the same sequence (page 5, lines 3-8, Figure 3B).

Claim 21 depends from claim 5 and further indicates that the peptide linker is between 8 and 25 amino acids in length (page 11, lines 21-25; Figure 4B).

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A. Whether claims 5, 6, 20 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Pomerantz (1998) *Biochemistry* 37(4):965-970 (hereinafter "Pomerantz") in view of Krylov et al. (1994) *EMBO J.* 13(12):2849-2861 (hereinafter "Krylov").

ARGUMENTS

A. The claims are non-obvious over the cited references

Claims 5, 6, 20, and 21 were rejected as allegedly obvious over Pomerantz and Krylov. (Final Office Action, pages 2 and 9 and page 2 of Advisory Action indicating that the claims remained rejected for the reasons set forth in the Final Office Action). Pomerantz was cited for allegedly disclosing a zinc finger protein fused to a naturally occurring dimerization domain extracted from the GAL4 protein and for suggesting the use of non-naturally occurring dimerization domains. *Id.* Krylov, reference 20 of

Pomerantz, was cited for allegedly demonstrating that non-naturally occurring peptide linkers could be utilized to complex zinc finger proteins. *Id.*

In response to Appellants' arguments that there is no combination of the references that teaches the claimed complexes, namely complexes in which zinc finger proteins are joined to each other by non-naturally occurring peptide linkers of less than 30 amino acids, it was again asserted that Pomerantz teaches zinc finger proteins with short peptide linkers for covalent linkages, which makes it obvious (predictable) to use such linkers for non-covalent linkages and that the "art appreciates the use or teaches the conventionality of a short length peptide linker to link two proteins whether by covalent or non-covalent linkage." (Final Office Action, page 5). In addition, it was again alleged that the paragraph bridging pages 965-966 of Pomerantz suggests fusion of heterologous modules with a short peptide linker. (Final Office Action, pages 6-7).

With regard to Krylov's failure to teach dimerization domains under 30 amino acids, the Examiner asserted that "there is nothing in the specification that demonstrates that the claimed 30-residue peptide produces new or unexpected results from that of Krylov, 32-residue." (Final Office Action, pages 10-11, citing *Merck v. Biocraft* and *KSR v. Teleflex*).

1. The references do not teach or suggest the claimed elements

In order to establish obviousness of a claimed invention, all the features of the claims must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). Thus, in the case on appeal, Pomerantz and Krylov must teach or suggest a complex comprising two fusion proteins in which each fusion protein comprises a zinc finger protein and a 30 amino acid or less non-naturally occurring peptide linker that forms a dimer with the corresponding non-naturally occurring peptide linker on a separate fusion protein. In addition, each zinc finger protein must bind to DNA in a sequence-specific manner.

For the reasons of record, there is no combination of the cited references that teaches or suggests the claimed elements. As acknowledged by the Examiner, Pomerantz is silent as to complexes including a non-naturally occurring peptide linker of the claimed

length. Rather, Pomerantz teaches only dimers formed using the naturally occurring GAL4 dimerization domain (Pomerantz, Abstract):

> Computer modeling was used to design a dimeric zinc finger protein, ZFGD1, containing zinc fingers 1 and 2 from Zif268 and a portion of the dimerization domain of GAL4.

The GAL4 dimerization domain used in Pomerantz is fully 50 amino acids in length (see, Figure 1 of Pomerantz). Notably, this does not include a 9 amino acid peptide linker used to join the zinc fingers to the dimerization domain. Thus, Pomerantz, at best, discloses modifying an almost 60 amino acid dimerization domain and linker, which is twice as long as in the claimed complexes. Simply put, Pomerantz does not teach or suggest non-naturally occurring peptide linkers of the claimed lengths.

Appellants also strongly traverse the allegation that Pomerantz suggests using Krylov's non-naturally occurring peptide linkers to join zinc finger proteins. (See, Final Office Action, citing page 970 of Pomerantz, which includes reference to Krylov (ref. 20), left column, emphasis added):

> As demonstrated by many studies, the coiled-coil interaction motif offers the potential to modify the dimerization domain to increase dimerization affinity or to specifically promoter heterodimer formation (see refs. 19 and 20 for examples). Obligate heterodimer variants of the zinc-finger-GAL4 fusion might be constructed to ensure that the proteins could only bind to the heterodimer site. It also seems plausible that adjustments to the linker region may give further improvements in the affinity and specificity of the zinc finger-GAL4 proteins.

In fact, in this passage Pomerantz is referring, at best, to alteration of the coiled-coil GAL-4 domain (or linker region). *See, e.g.,* page 966, left column of Pomerantz, emphasis added:

> The GAL4 domain was chosen because structural information is available for this domain (18) and because it contains a coiled-coil motif, a simple, well-understood structure that can be further modified for design purposes.

Thus, Pomerantz references Krylov as an example of how non-GAL-4 coiled-coil dimerization domains could possibly be altered. This is in no way a suggestion to take Krylov's leucine zipper from its natural context and use it to join zinc finger proteins. Rather, Pomerantz speculates that GAL4 (or the linker region) could possibly be altered. Not only is this pure speculation and not a demonstration of anything regarding altered GAL-4, it remains the case that under any of Pomerantz's unsupported theoretical constructs, the GAL-4 dimerization domain remains much longer than 30 amino acids in length. Thus, Pomerantz does not teach or suggest the claimed complexes.

Krylov fails to supply what is missing from Pomerantz. Krylov's dimerization domain is altered in its natural context of the leucine zipper protein as a whole. (Krylov, left column of page 2850 to left column of page 2851 and Fig. 1):

> The protein sequence of the first four leucine zipper heptads of the host or parent protein, the bZIP protein VBP (Iyer et al., 1991) is presented in Figure 1B. ....
>
> The lower section of Figure 1B presents the nomenclature used to describe our various mutant proteins.

There is nothing in Krylov (or Pomerantz) that teaches this domain could be isolated from its natural context and used to dimerize zinc finger proteins. Krylov is silent in this regard and, as noted above, Pomerantz references Krylov as an example of how GAL-4 could be altered, not to suggest that Krylov's domain could somehow be isolated and used with zinc finger proteins.

Furthermore, as noted repeatedly on the record and herein, like Pomerantz's GAL-4 domain, Krylov's coiled-coil domain is still more than 30 amino acids in length. In fact, Krylov's dimerization domains are at least 37 amino acids in length (4 heptads and 3 N-terminus amino acids and 2 C-terminal amino acids). *See,* Figure 1B of Krylov. Even if only the heptads (32 amino acids) were isolated from Krylov, Pomerantz teaches that a separate linker would be needed to join the dimerization domain to the zinc fingers. *See,* Figure 1 of Pomerantz. Accordingly, once linked to the zinc fingers, even the shortest possible heptad repeat of Krylov would still be over 40 amino acids.

Finally, Appellants also reiterate that the claims require that the peptide linker be non-naturally occurring. As clearly defined in the as-filed specification, a non-naturally occurring peptide linker is one that lacks significant sequence identity with a naturally occurring peptide (see, e.g., page 3, lines 22-24 of the as-filed specification):

> The invention provides non-naturally occurring dimerizing peptides. Some such peptides are homo-dimerizing peptides. Such peptide typically lack significant sequence identity with a naturally occurring peptide.

As both Pomerantz and Krylov are referring, at best, to alteration of very few amino acids of a naturally occurring dimerizing domain, it is clear that their domains will exhibit significant sequence identity with a naturally occurring peptide and, as such, are unlike the claimed molecules.

In sum, neither Pomerantz nor Krylov teach or suggest a non-naturally occurring peptide linker of 30 amino acids or less, as claimed. Therefore, there is no combination of Pomerantz and Krylov that teaches all the claimed elements and, on this basis alone, the rejection cannot stand.

2. The proposed combination does not give a predictable outcome

The obviousness rejection is also improper because the proposed combination of using Krylov's altered leucine zipper dimerization domains, shortening these domains to less than 30 amino acids and using these truncated domains in place of Pomerantz's GAL4 dimerization domains is not predictable from the teachings of the references and/or state of the art. Indeed, as the Supreme Court in *KSR v. Teleflex,* 550 U.S. 398, 127 S. Ct. 1727, 82 U.S.P.Q.2d 1385 (2007) reiterated, an obviousness inquiry is fact-dependent and that "a patent composed of several elements is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art." *KSR*, 82 UPSQ2d at 1389. Instead, the combination of elements must result in a predictable outcome (see, Examination Guidelines for Determining Obviousness Under 35 U.S.C. 103 in view of the Supreme Court Decision in *KSR International Co. v. Teleflex Inc.*, Fed. Reg. Vol. 72, No. 195, October 10, 2007, emphasis added):

> The rationale to support a conclusion that the claim would have been obvious is that all the claimed elements were known in the prior art and one skilled in the art could have combined the elements as claimed by known methods with no change in their respective functions, and the combination would have yielded nothing more than predictable results to one or ordinary skill in the art at the time of the invention.

It is also axiomatic that an obviousness rejection is improper where the proposed modification would destroy the intended function of the reference (see, e.g. *In re Fritch* 23 USPQ2d 1780, 1783, n.12 (Fed. Cir. 1992) and *In re Ratti* 123 USPQ 349, 352 (CCPA 1979)):

> A proposed modification [is] inappropriate for an obviousness inquiry when the modification render[s] the prior art reference inoperable for its intended purpose.
>
> [I]t would require a substantial reconstruction and redesign of the elements shown in [a cited reference] as well as a change in the basic principles under which [that reference's] construction was designed to operate.

As a threshold matter, Applicants again note that Krylov's dimerization domains, even in isolation, are longer than the maximum length recited in the claims. There is absolutely no evidence of record establishing that removing amino acids from the at least 32-amino acid heptads disclosed in Krylov (such that the dimerization domain contains less than 30 amino acids) is predictable. Indeed, the skilled artisan would recognize that isolating only the heptads and removing at least two amino acids from these heptads would likely destroy the intended dimerization function of Krylov's domains.

It is also unpredictable from the references that non-naturally occurring peptide linkers as claimed would dimerize in the context of zinc finger fusion proteins. In particular, Krylov fails to teach anything about predictability of dimerizing zinc finger proteins. For its part, Pomerantz clearly teaches that dimerization is only predictable when using known dimerization domains, which Pomerantz and Krylov both teach are longer than 30 amino acids in length (Pomerantz, page 966, emphasis added; also cited on page 6 of the Final Office Action):

> Dimer formation, frequently employed by **natural** DNA-binding proteins to enhance the affinity and specificity of recognition, provides **another** attractive design strategy. The capacity for homo- and heterodimerization offers several potential advantages for DNA binding proteins. ... Design strategies that employ dimerization also **may provide a useful alternative to the covalent linkage of multiple DNA-binding domains**. Large covalent assemblies might have higher absolute affinity for non-specific DNA sites and might become kinetically trapped at inappropriate sites in the genome. Dimerization provides **an alternative way** of bringing multiple domains together as a functional unit.

Therefore, contrary to the Examiner's assertion that covalent linkage with a short linker renders non-covalent linkage via dimerization domains predictable, Pomerantz is clear that any predictability in linking via dimerization lies in the use of naturally occurring dimerization domains of at least 50 amino acid residues. Pomerantz doesn't teach that shorter, non-naturally occurring linkers can be used for dimerization, as would be required to show obviousness. This reference only teaches that short (covalent) linkers can be used to fuse zinc finger proteins and that, as an alternative, the long, naturally occurring GAL-4 dimerization domain can be used to join zinc finger proteins.

Furthermore, as noted above, Pomerantz teaches nothing about the "predictability" of using Krylov's altered dimerizing domains in the context of zinc fingers. Rather, Pomerantz teaches that Krylov can be used as a model for further modifications to the GAL-4 domain used to dimerize zinc fingers might possibly be modified (See, Pomerantz, page 970, left column, emphasis added):

> As demonstrated by many studies, the coiled-coil interaction motif offers **the potential** to modify the dimerization domain to increase dimerization affinity or to specifically promoter heterodimer formation (see refs. 19 and 20 for examples). Obligate heterodimer variants of the zinc-finger-GAL4 fusion **might** be constructed to ensure that the proteins could only bind to the heterodimer site. **It also seems plausible** that adjustments to the linker region may give further improvements in the affinity and specificity of the zinc finger-GAL4 proteins.

Offering the "potential" or indicating that something "might be" constructed or "seems plausible" does not in any way establish that it was a predictable use of known elements.

Thus, the references do not teach complexes comprising non-naturally occurring peptides 30 or fewer amino acids in length for joining zinc finger proteins. Moreover, the evidence of record establishes that it was completely unpredictable to take altered dimerization domains out of their natural context, shorten them to 30 amino acids or less and then use them in the context of zinc finger proteins. The Examiner may wish it to be otherwise (or know it to be otherwise now based on the instant specification), but wishing and hindsight reconstruction cannot support an obviousness rejection.

For the reasons of record and as set forth above, the rejection to the claims on appeal cannot be sustained.

CONCLUSION

For the reasons stated above, Appellants respectfully submit that the pending claims are patentable.

Respectfully submitted,

Date: September 28, 2009

By: 

Dafna S. Pasternak
Registration No. 41,411
Attorney for Appellants

ROBINS & PASTERNAK LLP
1731 Embarcadero Road, Suite 230
Palo Alto, CA 94303
Telephone: (650) 493-3400
Facsimile: (650) 493-3440

CLAIMS APPENDIX

5. A zinc finger complex, comprising two or more fusion proteins, each fusion protein comprising a zinc finger protein that binds to DNA in a sequence-specific manner and a peptide linker, wherein the zinc finger proteins of the fusion proteins are joined to each other by specific binding of the peptide linkers, and wherein the peptide linkers are non-naturally occurring peptide linkers of 30 amino acids or less in length.

6. The zinc finger complex of claim 5, wherein the peptide linker of each fusion protein is the same.

20. The zinc finger complex of claim 5, wherein the zinc finger protein of each fusion protein has the same sequence.

21. The zinc finger complex of claim 5, wherein the peptide linker is between 8 and 25 amino acids in length.

EVIDENCE APPENDIX

No documents are attached to this appendix.

RELATED PROCEEDINGS APPENDIX

As noted above on page 2 of this Appeal Brief, a Board Decision on Appeal in this case was mailed on May 30, 2007 and is attached hereto.

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Ex parte BRYAN S. WANG, and CARL O. PABO

Appeal No. 2006-3085
Application No. 09/636,243

ON BRIEF

Before ADAMS, MILLS, and LEBOVITZ, Administrative Patent Judges.

LEBOVITZ, Administrative Patent Judge.

DECISION ON APPEAL

This appeal involves claims to zinc finger complexes of two or more fusion proteins, each fusion protein comprising a zinc finger joined to a peptide linker. The Examiner has rejected the claims as lacking written description, indefinite, and anticipated. We have jurisdiction under 35 U.S.C. § 6(b). We reverse the rejections, but set forth a new ground of rejection in which we find all claims obvious in view of prior art.

Background

Zinc finger proteins are DNA-binding proteins. Specification, page 1, lines 11-12. A single finger domain is about 30 amino acids long and recognizes a specific sequence

of nucleotides. Id., page 1, lines 15-16 and 23-30; page 7, lines 18-20. Protein engineering has been utilized to generate zinc fingers with novel DNA sequence specificities. Id., page 2, line 3-page 3, line 16. Because of their DNA-binding properties, zinc finger proteins have been adapted to a variety of different applications, including for sequence detection of target nucleic acid in a sample (id., page 27, lines 6-7) and to regulate gene expression (id., page 27, lines 27-32).

The instant application describes zinc finger proteins that have been appended with peptide sequences which are able to interact and bind to each other. The peptide sequences facilitate the formation of zinc finger protein complexes in which the zinc finger proteins are held together by interactions between the peptides and other contacts the peptides make. Id., page 11, lines 26-34.

Discussion

Claim construction

Claims 5, 6, and 20 are appealed. These are the only pending claims in the application. Claim 5 is the only independent claim and it reads as follows:

> 5. A zinc finger complex, comprising two or more fusion proteins, each fusion protein comprising a zinc finger protein and a peptide linker, wherein the fusion proteins are joined to each other by specific binding of the peptide linkers, and wherein the peptide linkers are non-naturally occurring peptides.

The claim is drawn to a complex which comprises at least two fusion proteins which are "joined to each other by specific binding of the peptide linkers." To understand this structure, we must first look at the construction of the fusion protein.

A fusion protein, according to the claim, comprises a zinc finger protein and a peptide linker. A "zinc finger protein" is defined in the Specification to be a protein which "binds DNA in a sequence-specific manner." Specification, page 7, lines 18-20. A zinc finger protein can contain from one to thirty-seven individual fingers, each finger of which binds to a defined subsite within the target DNA site. Id., page 14, lines 12-17. A zinc finger protein can be engineered to vary the order and nucleotide sequence specificity of its finger components. Id., page 14, line 22-page 15, line 15; page 7, lines 22-27.

A "peptide linker" is not expressly defined, but in the context of the Specification would be understood by the skilled artisan. The "Summary of the Claimed Invention" refers to both "dimerizing peptides" and "peptide linkers." Id., page 3, lines 22-25 and 26-31, respectively.

The dimerizing peptides "mediate association" of the zinc fingers to which they are attached. Id., page 11, lines 8-14. According to the Specification, a phage display method can be utilized to select the dimerizing peptides. Id., page 11, lines 15-22. In the description of this method, it is stated that "phage display is used for selection of linkers." (Emphasis added.) Id., page 19, line 22. "The method involves the generation of diverse libraries of peptides, typically linked to the same zinc finger protein, followed by affinity selection for phage bearing peptides with dimerizing activity." (Emphasis added.) Id., page 19, lines 22-25. In this context, the skilled artisan would recognize that the peptides which are "linked" to the zinc finger proteins are the "peptide linkers" recited in the claims. When the peptide linkers are selected for their ability to associate with each other (i.e., dimerize), they can also be referred to as "dimerizing peptides."

The limitation in claim 5 that the "fusion proteins are joined to each other by specific binding of the peptide linkers" requires that the peptide linkers attach to each other (i.e., "bind"), the same activity defined for the dimerizing peptides. As the "peptide linkers" have the same activity as the dimerizing peptides and also the same physical association with zinc fingers, we consider the two to be equivalent for the purposes of claim 5.

On page 15 of the Specification, the phrase "peptide linkers" is used to describe peptides which covalently join portions of the zinc fingers together. Not until the peptide linkers are selected for dimerizing activity would they be characterized as dimerizing peptides or the peptide linkers of claim 5 which join the fusion proteins together by "specific binding." Thus, it is clear in the context of the Specification that the claimed peptide linkers are peptides which have been selected for their ability to associate the fusion proteins by specific binding (i.e., homo- or hetero-dimerization as described in the Specification on page 11, lines 8-14).

The peptide linkers are characterized by the claim as "non-naturally occurring peptides." The term "non-naturally occurring" is defined in the Specification to refer to "objects and sequences not found in nature." Id., page 8, lines 3-4. Typical and preferred non-naturally occurring sequences are described. The examples in the Specification describe the selection of peptides with dimerizing activity. See, e.g., id., page 19, lines 24-25. The dimerizing peptides are present in fusion proteins where they are linked to zinc finger proteins. Id., page 19, lines 32-34; page 20, lines 1-7; page 28, lines 15-18. In these cases, the term "peptide" is being used as shorthand for the amino acid sequence which is contained within the fusion protein. It is the amino acid

sequence of the dimerizing peptide which possesses the claimed "specific binding" activity enabling them to dimerize the fusion proteins. Accordingly, we interpret the phrase "non-naturally occurring peptides" to mean that the peptide linker sequence (i.e., the dimerizing "peptide linker") is not found in nature.

Consistent with this construction, the Specification states that the selection of "peptides" having "novel" dimerization motifs makes it less likely that they will react "with natural dimerization interfaces presented by proteins in the cell." Id., page 38, lines 30-34. That is, a naturally occurring sequence (i.e., "peptide linker" of claim 5) within the fusion protein is disfavored because it would be more apt to specifically bind to its natural dimerizing partner within a normal cellular protein than a sequence which is artificial ("non-naturally occurring"). In this context, both the peptide and the "natural dimerization interface" are amino acid sequences.

The fusion proteins are joined "by specific binding of the peptide linkers." The Specification does not define "specific binding" in the context of the peptide linkers, but it does state that the peptides "mediate association" of the fusion proteins and that the "proteins ... bind to each other via the dimerizing peptides." Id., page 11, lines 8-14 and page 13, lines 5-6, respectively. The association/binding is characterized as involving "supporting interactions [that] can include contacts between the peptides and/or contacts between a peptide and another region of the protein." Id., page 11, lines 30-34. It is not a covalent interaction involving covalent bonds.

In sum, the claim is drawn to a complex of at least two fusion proteins, where each protein comprises a zinger finger which is fused to a peptide sequence. The

peptide sequence is not found in nature (i.e., non-naturally occurring). Contacts between the peptide sequences in each fusion protein hold the complex together.

Rejection under § 112, first paragraph for lack of written description

Claims 5, 6, and 20 stand rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the Specification. The Examiner stated that the "as-filed Specification does not describe a claimed zinc finger complex comprising two or more fusion proteins linked by peptide linkers that are non-naturally occurring peptides." Answer, page 4.

"To fulfill the written description requirement, the patent Specification 'must clearly allow persons of ordinary skill in the art to recognize that [the inventor] invented what is claimed.' In re Gosteli, 872 F.2d 1008, 1012, 10 USPQ2d 1614, 1618 (Fed. Cir. 1989). An applicant complies with the written description requirement 'by describing the invention, with all its claimed limitations.' Lockwood v. American Airlines, Inc., 107 F.3d 1565, 1572, 41 USPQ2d 1961, 1966 (1997)." Gentry Gallery v. The Berkline Corp., 45 USPQ2d 1498, 1502-1503 (Fed. Cir. 1998).

Claim 5, as originally filed, reads as follows:

> 5. A zinc finger complex, comprising a first fusion protein comprising a first zinc finger protein and a first peptide linker and a second fusion protein comprising a second zinc finger protein and a second peptide linker, wherein the first and second fusion proteins are complexed by specific binding of the first and second peptide linkers, and wherein the first and second peptide linkers are nonnaturally occurring peptides.

Comparing amended claim 5, which is now on appeal, to original claim 5, we find almost all its key claimed limitations, including that the fusion proteins each comprising

a "zinc finger protein" and "peptide linker", that the fusion proteins are joined ("complexed" in the original) by "specific binding", and that the peptide linkers are "non-naturally occurring peptides." The only substantial disparity is that amended claim 5 recites that the complex can comprise "two or more fusion protein," while the original did not contain this express limitation. The Examiner argued that the amended claim lacked support in the disclosure, stating that the disclosure "does not describe two linkers each linked to each ZIF [zinc finger protein] wherein the linkers specifically binds to each other to form a fusion of two zinc fingers as schematically depicted by appellants at page 9 of the 8/8/03 Brief." Answer, page 7. As we understand it, the Examiner's position appears to be that there is no written description for complexes which contain more than two zinc finger fusion proteins held together by specifically binding peptide linkers.

To resolve this issue, we turn to the written description. First, we address whether there is support in the Specification for the claimed subject matter of two fusion proteins held together by specific binding between peptide linkers. We discuss this issue in more detail because the Examiner appears to question whether this embodiment was described in the Specification.

As we have construed the claim, the recited "peptide linker" having "specific binding" activity is the same element referred to in the Specification as a "dimerizing peptide." The skilled worker would have gleaned this concept upon reading the Specification in its entirety. For example, in the "Detailed Description," pages 11-14 are devoted to a description of how peptides are selected from peptide libraries, particularly

random peptide libraries, to identify sequences capable of associating two molecules together (i.e., dimerizing).

The relationship between the peptide linkers utilized in the phage library and the dimerization peptides is also shown in Specification Fig. 3. Fig. 3A is a drawing of a DNA construct containing zinc fingers fused to random peptides ("peptide library"). Fig. 3B shows a dimer of fusion protein products of the DNA construct bound to target DNA. It would be evident that the peptide ("peptide library") linked to ZiF12 (a zinc finger protein) depicted in Fig. 3A, after it has been selected for the dimerizing activity, is equivalent to the "dimerizing peptides" of Fig. 3B

The examples in the Specification also provide support for the written description of the claimed subject matter. On page 32 ("Results"), it is stated that "[t]o select dimerization motifs, we attached random peptides to a DNA-binding domain and selected those fusion proteins that could bind more stably to a symmetric DNA site." The random 15- and 30-amino acid peptides were "expressed at the amino terminus of the first two zinc fingers of Zif268." Id. The zinc finger component mediated binding to target DNA. The binding of monomers – a single fusion protein containing a peptide and zinc finger – and dimers – two of the fusion proteins associated together by the peptide linker – to DNA were characterized. Id., Table 1; page 33, lines 25-page 34, line 7. These experiments clearly demonstrate Appellants' possession of the claimed zinc finger complexes containing two zinger fingers joined to each other by binding between the peptide linkers. We do not find these examples to be incongruent with the Specification's general description as the Examiner contends. Answer, page 9.

We also find support in the Specification for the limitation that the complex contains "two or more" fusion proteins. As pointed out by Appellants (Brief, page 9), it is stated on page 12 of the Specification that dimerizing peptides "are useful for mediating multimerization of zinc finger proteins." A multimer is defined as a "protein made up of more than one peptide chain,"[1] and would include dimers of two fusion proteins (Fig. 3) and trimers of three (Brief, page 9). Although the only examples are of dimers, the Specification refers to these as a "typical application" of the technology, but does not exclude other embodiments. Specification, page 12, lines 21-22. On page 15 of the Specification, it is expressly stated that "[t]wo or more zinc finger proteins can be linked either covalently or by dimerization." Id., page 15, lines 4-5. It is our view this provides explicit support for the claim language (added by amendment on December 27, 2002) "two or more fusion proteins."

For the foregoing reasons, it is our view that the Specification provides an adequate written description of the claimed subject matter. This rejection is reversed.

Rejection under § 112, second paragraph

Claims 5 and 20 stand rejected under 35 U.S.C. § 112, second paragraph, as failing to point out and distinctly claim "the subject matter which the applicant regards as his invention." The Examiner stated that "[i]t is not clear within the claimed context as to what constitutes a [sic] non-naturally-occurring peptide linkers." Answer, page 5. Appellants challenged the rejection, arguing that "non-naturally occurring" is clearly

[1] Zaid et al., Glossary of biotechnology and genetic engineering, FAO Research and Technology Paper, 158 (1999).

defined in the Specification to include only those sequences not found in nature. Brief, page 11.

A specification must conclude with claims "particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention." 35 U.S.C. § 112, second paragraph (2000). The purpose of §112, second paragraph, is to "reasonably apprise those skilled in the art of the scope of the invention." Miles Labs., Inc. v. Shandon, Inc., 997 F.2d 870, 875, 27 USPQ2d 1123, 1126 (Fed. Cir. 1993).

We agree with Appellants that "non-naturally occurring," which is recited in claim 5, would "reasonably apprise" the skilled artisan of the scope of the claimed subject matter. In particular, page 8, lines 3-4, of the Specification states that "non-naturally[ ] occurring is used to describe objects and sequences not found in nature." We find no merit in the Examiner's arguments characterizing the description of non-naturally occurring in the Specification as inconsistent with the "conventional wisdom of the art," and confusing in its reference to protein folding. Answer, page 12, paragraph 2. The statements referred to by the Examiner described "preferred," "typical," and "some" embodiments, which are examples of non-naturally occurring sequences.

For the reasons set forth above, we reverse this rejection.

Rejection under § 102(b)

Claims 5 and 20 stand rejected under 35 U.S.C. § 102(b) as anticipated by Pomerantz.[2]

Pomerantz describes the "design of a dimeric zinc finger protein, ZFGD1, containing zinc fingers 1 and 2 from Zif 268 and a portion of the dimerization domain of GAL4." Pomerantz, Abstract. The GAL4 dimerization domain is utilized to associate two chimeric proteins, each containing a zinc finger fused to the GAL4 domain, i.e., a zinc finger-GAL4 fusion. Id., page 966, column 2; Fig. 1. GAL4 is a naturally-occurring protein that contains a coiled-coil dimerization motif that mediates "protein-protein interaction." Id. Binding studies of the fusion protein established that ZFGD1 binds to DNA as a dimer. Id., page 967, column 2.

The Examiner argued that Pomerantz's disclosure of the zinger finger-GAL4 fusion protein anticipates claims 5 and 20.

> Pomerantz recites that a portion of Gal4 is used as dimerizing linker. This portion is shown at page 967, col. 1 under the heading section RESULTS i.e., the portion that binds to the 13-residue DNA substite [sic]. Read in the light of the specification definition of a non-naturally occurring peptide linkers e.g., less than 50% (amino acid) with natural sequences the GAL4 (41-100 residues) is less than 50% of the naturally occurring sequence of Gal.

Answer, paragraph spanning pages 13-14, emphasis removed.

Appellants contended that "[t]he claims on appeal clearly require linking of two or more proteins via non-naturally occurring peptides. In contrast, the dimerizing linker used by Pomerantz, namely amino acids 41 to 100 of GAL4, is clearly a naturally-occurring peptide sequence, inasmuch as it is part of the naturally occurring

[2] Pomerantz et al., (Pomerantz), Biochemistry, 37(4):965-70 (1998).

GAL4 protein." Supplemental Brief, page 8, paragraph 2. We agree with Appellants' position.

Anticipation requires a showing that each element of the claim is identifiable in a single reference. Perricone v. Medicis Pharm. Corp., 432 F.3d 1368, 1375, 77 USPQ2d 1321, 1325 (Fed. Cir. 2005). The GAL4 domain utilized in Pomerantz's fusion protein is a naturally-occurring sequence obtained from the naturally-occurring GAL4 protein. Having interpreted the claims to require that the peptide linker with specific binding ("dimerizing") activity is a non-naturally occurring sequence, we are compelled to conclude that Pomerantz does not teach each and every element of the claimed subject matter. Consequently, we find that the Examiner has failed to establish adequate evidence of prima facie anticipation. This rejection is reversed.

New Grounds of Rejection

Under the provisions of 37 CFR § 41.50(b), we enter the following new grounds of rejection.

Claims 5, 6, and 20 are rejected under 35 U.S.C. § 103(a) as unpatentable over Pomerantz in view of Krylov.[3]

The Pomerantz publication has been described above for its disclosure of a zinc finger fused to the naturally occurring dimerization domain extracted from the GAL4 protein. Pomerantz's fusion protein differs from the fusion protein contained in the zinc finger complex of claim 5 by having a naturally occurring dimerization domain, instead

[3] Krylov et al. (Krylov), The EMBO Journal, 13(12):2849-61 (1994).

of the non-naturally occurring sequence ("peptide linker") required by the claim. However, in addition to its disclosure of a zinc finger-GAL4 fusion protein, Pomerantz suggests that other dimerizing domains may be appended to the zinc finger.

> The dimerization interface also provides opportunities for further elaboration and optimization. As demonstrated by many studies, the coiled-coil interaction motif offers the potential to modify the dimerization domain to increase dimerization affinity or to specifically promote heterodimer formation (see refs 19 and 20 for examples).
>
> ....
>
> Dimer contacts of modest affinity may allow self-assembly at the appropriate binding site and thereby reduce the risks of nonspecific (kinetic) trapping that may occur with large covalently linked sets of fingers. Cooperative binding, by giving a more dramatic concentration dependence, may also allow more precise on/off switching in targeted gene regulation.

Pomerantz, page 970, column 1.

A prima facie case of obviousness requires evidence that the prior art disclosed or suggested all of the elements of the claimed invention, and that those skilled in the art would have been motivated to combine those elements with a reasonable expectation of success. See In re Wilson, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970); In re Vaeck, 947 F.2d 488, 493, 20 USPQ2d 1438, 1443 (Fed. Cir. 1991).

Here, Pomerantz does not teach that the zinger finger protein is attached to a non-naturally occurring dimerization domain as required by claim 5, but suggests that such domains be used "to increase dimerization affinity," "to specifically promote heterodimer formation," and "allow more precise on/off switching in targeted gene regulation." Pomerantz, page 970, column 1. This provides the motivation which would have led one of ordinary skill in the art to have replaced the GAL4 dimerization domain with non-naturally occurring sequences having the requisite dimerizing activity.

Pomerantz points the skilled artisan directly to prior art publications that teach modified dimerization domains. Such domains are non-naturally occurring and "join each other by specific binding," meeting the requirements of the claimed "peptide linkers." See claim 5. In particular, reference 19 (hereinafter "Krylov"), cited by Pomerantz for its studies of the coiled-coil interaction motif, describes "protein design rules that can be used to modify leucine zipper-containing proteins to possess novel dimerization properties." Krylov, page 2850, column 1. "33 different leucine zipper proteins containing 27 different systematic combinations of amino acids" were produced. Id., page 2856, column 2 ("Discussion"). See also Fig. 1B for a list of exemplary "mutant proteins." Id., page 2850, column 2. The mutant proteins were mixed together under conditions which facilitated dimer formation. By measuring the stability of the dimers formed (id., page 2852-53, "Thermodynamic stability"), Krylov was able to demonstrate that certain modified dimers had increased stability and specificity as compared to the unmodified form. ("Novel heterologous interactions regulate dimerization specificity. ... In the second mixing experiment, the stability of the heterodimer is calculated to be greater than the average of the two homodimer stabilities, thus favoring the formation of heterodimers." Id., page 2856, columns 1-2.) Thus, the element missing from Pomerantz – non-naturally occurring peptide linkers – is supplied by Krylov. The skilled worker would have had a reasonable expectation that Krylov's domains could be utilized to complex zinc fingers to which they are attached in view of Krylov's success in not only modifying their binding activity, but in making it stronger (i.e., more stable).

Krylov also teaches dimerization domains having the same sequence, meeting the limitations of clam 6. See e.g., id., page 2856, column 1, describing homo- and heterodimers, where the homodimers have "the same sequence."

Pomerantz describes dimers between ZFGD1 fusion protein, where each fusion contains the same zinc finger. Pomerantz, Abstract ("a dimeric zinc finger protein, ZFGD1"). This meets the requirements of claim 20.

In sum, we find that Pomerantz and Krylov disclose all elements of the subject matter recited in claims 5, 6, and 20. For the reasons discussed above, the skilled worker would have considered these claims obvious in view of Pomerantz's express suggestion to combine its teaching with Krylov (i.e., reference 19), and Krylov's disclosure that would have led the skilled worker to reasonably expect that the combination would work.

Summary

The rejections of claims 5, 6, and 20 are reversed. A new ground of a rejection has been entered under § 103 for claims 5, 6, and 20.

Time Period for Response

This decision contains a new ground of rejection pursuant to 37 CFR § 41.50(b) (effective September 13, 2004, 69 Fed. Reg. 49960 (August 12, 2004), 1286 Off. Gaz. Pat. Office 21 (September 7, 2004)). 37 CFR § 41.50(b) provides "[a] new ground of rejection pursuant to this paragraph shall not be considered final for judicial review."

37 CFR § 41.50(b) also provides that the appellant, WITHIN TWO MONTHS FROM THE DATE OF THE DECISION, must exercise one of the following two options with respect to the new ground of rejection to avoid termination of the appeal as to the rejected claims:

> (1) *Reopen prosecution.* Submit an appropriate amendment of the claims so rejected or new evidence relating to the claims so rejected, or both, and have the matter reconsidered by the examiner, in which event the proceeding will be remanded to the examiner. . . .
>
> (2) *Request rehearing.* Request that the proceeding be reheard under § 41.52 by the Board upon the same record. . . .

REVERSED, 37 CFR 41.50(b)

Donald E. Adams)
Administrative Patent Judge)
)
)
Demetra J. Mills) BOARD OF PATENT
Demetra J. Mills)
Administrative Patent Judge) APPEALS AND
)
) INTERFERENCES
)
Richard M. Lebovitz)
Administrative Patent Judge)

RL/dm

COOLEY GODWARD KRONISH LLP
3000 EL CAMINO REAL
5 PALO ALTO SQUARE
PALO ALTO CA 94306

Notice of References Cited	Application/Control No. 09/636,243	Applicant(s)/Patent Under Reexamination Appeal No. 2006-3085	
	Examiner BPAI	Art Unit 1600	Page of

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Zaid, Glossary of biotechnology and genetic engineering, FAO Research & Technology Paper, 158 (1999)
	V	Krylov, The EMBO Journal, 13(12):2849-61 (1994)
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

FAO RESEARCH AND TECHNOLOGY PAPER No. 7

Glossary of biotechnology and genetic engineering

Note, in 2002, FAO published a revised,augmented version of this glossary. It is available at http://www.fao.org/DOCREP/004/Y2775E/Y2775E00.HTM or, as a searchable database, at http://www.fao.org/biotech/index_glossary.asp.

Table of contents

A. Zaid
H.G. Hughes
E. Porceddu
F. Nicholas

ISBN: 92-5-104369-8

ISSN: 1020-0541

Food and Agriculture Organization of the United Nations

Glossary of biotechnology and genetic engineering

A. Zaid
H.G. Hughes
E. Porceddu
F. Nicholas

FAO
RESEARCH
AND
TECHNOLOGY
PAPER
7

Food
and
Agriculture
Organization
of the
United
Nations

Rome, 1999

morphogenesis The development, through growth and differentiation, of form and structure in an organism.

morphogenic response The effect on the developmental history of a plant or its parts exposed to a given set of growth conditions or to a change in the environment.

morphology (Gr. *morphe*, form + *logos*, discourse) 1.The science of studying form and its development.

2. General: Shape, form, external structure or arrangement.

mosaic An organism or part of an organism that is composed of cells with different origin.

mother plant *See* donor plant.

movable genetic element *See* transposon.

mRNA; messenger RNA The RNA transcript of a protein-encoding gene. The information encoded in the mRNA molecule is translated into a polypeptide of specific amino acid sequence by the ribosomes. In eukaryotes, mRNAs transfer genetic information from the DNA to ribosomes, where it is translated into protein.

MRUs Minimum recognition units. *See* dabs.

mtDNA *See* mitochondrial DNA.

multi-copy Describing plasmids which replicate to produce many plasmid molecules per host genome, e.g., pBR322 is a multi-copy plasmid, there are usually 50 pBR322 molecules (or copies) per *E. coli* genome.

multigene family A group of genes that are similar in nucleotide sequence or that produce polypeptides with similar amino acid sequences.

multigenic Controlled by several genes, as opposed to monogenic.

multi-locus probe A probe that hybridizes to a number of different sites in the genome of an organism. *See* probe.

multimer; multimeric A protein made up of more than one peptide chain.

multiple alleles The existence of more than two alleles at a locus in a population.

multiple cloning site *See* polylinker.

multiple drop array (MDA) *See* microdroplet array.

multiple ovulation and embryo transfer (MOET) A technology by which a single female that usually produces only one or two offspring can produce a litter of offspring. Involves stimulation of a female to shed large numbers of ova; natural mating or

The EMBO Journal vol.13 no.12 pp.2849 – 2861, 1994

A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions

Dmitry Krylov, Irina Mikhailenko and Charles Vinson

Building 37, Room 4D06, Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Communicated by U.Schibler

The leucine zipper is a dimeric coiled-coil protein structure composed of two amphipathic α-helices with the hydrophobic surfaces interacting to create the dimer interface. This structure has been found to mediate the dimerization of two abundant classes of DNA binding proteins: the bZIP and bHLH-Zip proteins. Several workers have reported that amino acids in the e and g positions of the coiled coil can modulate dimerization stability and specificity. Using the bZIP protein VBP as a host molecule, we report a thermodynamic scale ($\Delta\Delta G$) for 27 interhelical interactions in 35 proteins between amino acids in the g and the following e positions (g⟷e') of a leucine zipper coiled coil. We have examined the four commonly occurring amino acids in the e and g positions of bZIP proteins, lysine (K), arginine (R), glutamine (Q), glutamic acid (E), as well as the only other remaining charged amino acid aspartic acid (D), and finally alanine (A) as a reference amino acid. These results indicate that E⟷R is the most stable interhelical pair, being 0.35 kcal/mol more stable than E⟷K. A thermodynamic cycle analysis shows that the E⟷R pair is 1.33 kcal/mol more stable than A⟷A with −1.14 kcal/mol of coupling energy ($\Delta\Delta G_{int}$) coming from the interaction of E with R. The E⟷K coupling energy is only −0.14 kcal/mol. E interacts with more specificity than Q. The R⟷R pair is less stable than the K⟷K by 0.24 kcal/mol. R interacts with more specificity than K. Q forms more stable pairs with the basic amino acids K and R rather than with E. Changing amino acids in the e position to A creates bZIP proteins that form tetramers.

Key words: coiled coil/dimerization specificity/leucine zipper/thermodynamic cycle

Introduction

The coiled coil is a helical protein motif that forms a variety of oligomers: dimers, trimers and tetramers (Cohen and Parry, 1990; O'Shea *et al.*, 1991; Alberti *et al.*, 1993; Harbury *et al.*, 1993; Lovejoy *et al.*, 1993). Dimeric coiled coils have a seven-residue repeat of hydrophobic and hydrophilic amino acids capable of forming an amphipathic α-helix (Figure 1A). In order to generate a repeating helical dimerization interface, the α-helix over-twists slightly, going from 3.6 to 3.5 amino acids/turn. This results in a repeating structural unit of two helical turns or seven amino acids (a heptad repeat). The **a** and **d** residues are hydrophobic and pack in a regular 'knobs and holes' pattern (Crick, 1953) along the dimerization interface. This creates the hydrophobic core that stabilizes the coiled coil and is critical for dimerization. The **e** and **g** positions which flank the dimerization interface contain a large number of charged amino acids and have been thought to interact electrostatically (Figure 1A) (Cohen and Parry, 1990; Alber, 1992; Baxevanis and Vinson, 1993). Recent work from Kim and Alber's groups (Harbury *et al.*, 1993) has shown that leucine in the **d** position is a critical determinant of the oligomerization properties of coiled coils; leucine favors dimers. This result suggests that the term leucine zipper is an appropriate nomenclature for parallel dimeric coiled coils (Landschultz *et al.*, 1988). The leucine zipper dimerization motif is critical for the functioning of two classes of DNA binding proteins: the bZIP (Vinson *et al.*, 1989) and bHLH-Zip proteins (Murre *et al.*, 1989).

The study of the essential structural elements that regulate leucine zipper dimerization stability and specificity has been facilitated by the fact that bZIP DNA binding is dependent on the correct dimerization of the leucine zipper structure. Several groups have shown that the hydrophobic core created by amino acids in the **a** and **d** position is critical for dimerization (Kouzarides and Ziff, 1988; Gentz *et al.*, 1989; Landschultz *et al.*, 1989; Turner and Tjian, 1989). As originally proposed by Landschultz *et al.* (1988), bZIP proteins not only homodimerize, but also heterodimerize via the leucine zipper structure (Hai *et al.*, 1989; Ivashkiv *et al.*, 1990; Roman *et al.*, 1990; Cao *et al.*, 1991; Hai and Curran, 1991; Williams *et al.*, 1991; Schindler *et al.*, 1992). However, only specific bZIP protein pairs can heterodimerize. Experiments using chimeric proteins and peptides indicate that all the structural information needed to regulate dimerization specificity is contained within the leucine zipper region (Agre *et al.*, 1989; Kouzarides and Ziff, 1989; O'Shea *et al.*, 1989).

The X-ray and NMR (Saudek *et al.*, 1991) structures of the GCN4 leucine zipper indicate that Crick's 'holes and knobs' description (Crick, 1953) of the dimer packing interface is a valid model. Amino acids in the **e** and **g** positions pack over the hydrophobic core created by the **a** and **d** positions, and possibly interact electrostatically. In the X-ray structure of the GCN4 leucine zipper (O'Shea *et al.*, 1991), GCN4 complexed to DNA (Ellenberger *et al.*, 1992; Konig and Richmond, 1993), and molecular dynamic calculations of the GCN4 zipper (Nilges and Brunger, 1991) interactions were observed between the **g** residue and the following **e** residue positioned five amino acids C-terminal on the opposite helix (denoted **e'**). Throughout this report, we refer to that interaction as the g⟷e' pair (when discussing a particular g⟷e' pair, the g and e' will be replaced by the relevant amino acid one-letter code). Several groups have implicated the amino acids in the **g** and **e** positions in the regulation of dimerization

specificity and stability (Nicklin and Casari, 1991; Schmidt-Dor *et al.*, 1991; Schuermann *et al.*, 1991; O'Shea *et al.*, 1992; Amati *et al.*, 1993; Loriaux *et al.*, 1993; Vinson *et al.*, 1993). Two groups (Hu *et al.*, 1993; Pu and Struhl, 1993) used a random mutagenesis approach to examine the significance of amino acids in the g and e position to dimer function and concluded that these positions are less important than the a and d positions for leucine zipper formation.

Work in this laboratory has shown that changing amino acids in the g and e positions of the homodimerizing bZIP protein C/EBP changes its dimerization properties (Vinson *et al.*, 1993). Proteins were generated that preferentially heterodimerized with wild-type C/EBP and are, therefore, potential dominant-negative proteins. Others were designed not to heterodimerize with C/EBP, but instead preferentially homodimerize. All these proteins were designed based on the assumption that E and R would create the most stabilizing g⟷e′ pair. In order to proceed further with leucine zipper design, we needed quantitative information on the relative strengths of different g⟷e′ pairs. The repeating nature of the leucine zipper structure suggests that design rules developed for a specific g⟷e′ pair within a particular heptad can be used for any heptad, irrespective of their amino acid content. We expected the interaction energies to be additive over the length of the dimerization interface, a proposal supported by data from previously designed proteins (Vinson *et al.*, 1993).

A desire to understand the energetic rules governing leucine zipper dimerization stability and specificity has motivated our studies to determine the energetic contribution of pairs of amino acids on opposite sides of the structure. In this report, we present thermodynamic measurements of the energetic contribution of 27 systematic g⟷e′ amino acid pairs to the thermal stability of the bZIP protein VBP (Iyer *et al.*, 1991), the chicken equivalent of the mammalian DBP (Mueller *et al.*, 1990). We have examined different combinations of six amino acids in the e and g positions: the four most common for these two positions, glutamic acid (E), glutamine (Q), arginine (R) and lysine (K), as well as aspartic acid (D) and alanine (A) (Vinson *et al.*, 1993). The dimerization stability of each new amino acid combination was determined using thermal melting monitored by circular dichroism (CD). These studies describe a thermodynamic scale for the stability of different g⟷e′ amino acid pairs. The results produce protein design rules that can be used to modify leucine zipper-containing proteins to possess novel dimerization properties.

Results

The host bZIP protein: VBP (the chicken DBP)

The protein sequence of the first four leucine zipper heptads of the host or parental protein, the bZIP protein VBP (Iyer *et al.*, 1991), is presented in Figure 1B. The g⟷e′ pairs immediately C-terminal to the DNA binding region are shown and numbered 1–4 depending on the heptad (defined here as **g,a,b,c,d,e,f**) within which they are found. The g position N-terminal of the first heptad interacts with DNA, as evidenced by the complex of GCN4 bZIP protein bound to DNA (Ellenberger *et al.*, 1992). Figure 1C presents a cartoon of a bZIP protein with the VBP amino acid sequence shown on a schematic of a coiled coil using the standard nomenclature for the seven unique amino acid positions

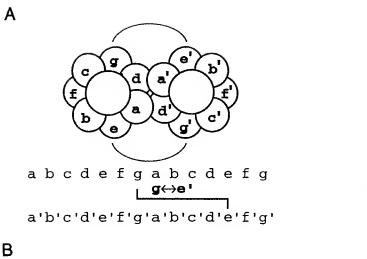

Fig. 1. (**A**) End view of a leucine zipper dimer looking from the N-terminus. The letters on the inside of each circle represent standard nomenclature for the seven amino acids found in unique positions in a coiled coil. Amino acids at the a and d positions create a hydrophobic core between the interacting helices. The interaction seen between amino acids in the g and subsequent e′ position seen in X-ray structures is noted as g⟷e′ pairs. (**B**) The amino acid sequence of the leucine zipper region of VBP, the chicken version of the mammalian DBP, is presented using the single-letter code. Below the VBP sequence is the nomenclature for the positions in a coiled coil. The sequence starts at the first 'leucine' position as defined previously (Vinson *et al.*, 1989) and is grouped into heptads (**g,a,b,c,d,e,f**). The leucine positions are italicized. The g to following e′ (g⟷e′) pairs are denoted by bars above the potentially interacting amino acids that are highlighted in bold type face. The heptads are numbered 1–4. Note that because of the 2-fold symmetry of the dimers, each heptad contains two g⟷e′ pairs. Amino acids in the g and e positions of representative mutant proteins are shown at the bottom of the figure to illustrate the nomenclature used. (**C**) The positions of the VBP leucine zipper amino acids seen on a schematic of a bZIP protein (Hu *et al.*, 1990) viewed from the side. Amino acids in the e and g position are shown in bold face. The heptad letter designations (**a,b,c,d,e,f,g**) are to the right of the figure. The supercoiling of the two helices is not depicted. To the left of the leucine zipper is the basic region of bZIP proteins with the DNA shown. The heptad N-terminal of heptad 1 has been shown to interact with DNA.

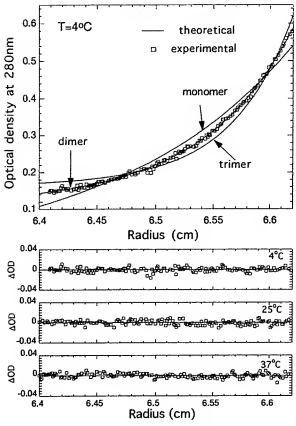

Fig. 2. Sedimentation equilibrium determination of the mol. wt of E·K$_{1234}$ C-S at 4°C. The sample was in 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 1 mM EDTA. Each sample was loaded at three concentrations (10, 20 and 40 μM) which have ODs of 0.1, 0.2 and 0.4 at 280 nm. Samples were spun at 25 000 r.p.m. for 24 h at 4°C. Theoretical curves for monomer, dimer or trimer are plotted as solid lines. The actual data are plotted as circles. These data clearly fit onto the dimer curve. The bottom panel shows the residual plots of fitting the experimental data to a monomer–dimer equilibrium model. No systematic error is evident.

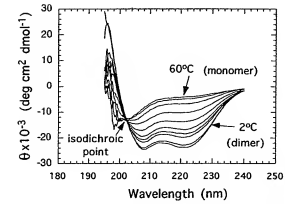

Fig. 3. Far-UV CD spectra of a 3.4 μM E·R sample as a function of temperature. The temperatures examined are from the highly ordered (dimer) to the random coil (monomer): 2, 6, 11, 20, 25, 30, 34, 40, 45, 45, 50 and 60°C. The minima at 208 and 222 nm are seen to disappear as the sample is heated. There is an isodichroic point at 202 nm. The sample was in 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 1 mM EDTA.

(Hodges *et al.*, 1972). We chose to use VBP as the host molecule for three reasons. First, it is the only known bZIP protein which has the potential for attractive g⟷e′ interactions at all four heptads, the most in any other protein is two. The most common attractive g⟷e′ pair, found in the second, third and fourth heptads of VBP, has an acidic amino acid in the g position and a basic amino acid in the following e position (Vinson *et al.*, 1993). The first heptad has the opposite charge configuration (Figure 1B). The second reason for choosing this protein is that 8 of 12 charged amino acids in the leucine zipper region are in the g or e positions rather than in the **b**, **c** or **f** positions, thus potentially simplifying the electrostatic interactions on the coiled-coil surface. This contrasts with C/EBP, for example, where only 4 of 13 charged amino acids are in the e or g positions. The final reason for using VBP is the ease of overexpression in *Escherichia coli*.

The lower section of Figure 1B presents the nomenclature used to describe our various mutant proteins. All proteins are defined by two letters separated by a large period. Each letter represents the standard code for an amino acid in the relevant g⟷e′ pair, the first letter being the amino acid in the g position and the second letter the amino acid in the following e position. All proteins are defined relative to the protein E·R, where the second, third and fourth g⟷e′ pairs are E⟷R, and the first pair is R⟷E. If a protein deviates from E·R, the g⟷e′ pair which differs is identified. Thus, for example E·K$_4$, which is the original VBP protein, has R⟷E in the first heptad, E⟷R in the second and third heptads, and E⟷K in the fourth heptad. Several proteins have the cysteine replaced by serine and are referred to by the note C-S.

Sedimentation equilibrium of host bZIP protein

Gel-shift mixing experiments indicate that wild-type VBP binds DNA as a dimer (Vinson *et al.*, 1993). Sedimentation equilibrium data show that E·K$_{1234}$ C-S behaves as a dimer at 4°C, even in the absence of DNA (Figure 2). At higher temperatures, 25 and 37°C, the sedimentation equilibrium data were fitted to a monomer–dimer equilibrium showing the temperature-induced dissociation of the dimer. Most of the mutant proteins to be discussed behaved as dimers at low temperatures.

Circular dichroism experiments of host bZIP protein: a simple two-state system

To determine whether or not helicity measured by CD would correlate with the extent of dimerization of the leucine zipper region observed by sedimentation equilibrium, we recorded far-UV CD spectra of a 3.4 μM solution of E·R as a function of temperature (Figure 3). The low-temperature spectra have minima at 208 and 222 nm, the signature for α-helices (Cooper and Woody, 1990). As the temperature is increased, these minima disappear, indicating the melting of the α-helical region. The presence of a clear isodichroic point (position common to all curves) is consistent with the two-state nature of the helix–random coil transition. Two-state melting was also observed recently for another bZIP protein, GCN4, in both CD and microcalorimetry studies (Thompson *et al.*, 1993). The thermal denaturation curves of our samples are reversible. Since VBP shows both reversible thermal

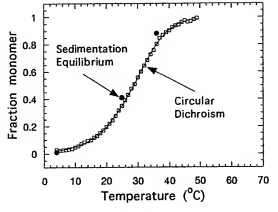

Fig. 4. The CD curve of $E{\cdot}K_{1234}$ C-S showing the fraction monomer as a function of temperature. The line through the data is a fitted curve as described in Materials and methods. Sedimentation equilibrium experiments were used to determine K_ds for the sample at 25 and 37°C. The K_ds obtained (2.4×10^{-5} M at 25°C and 5.5×10^{-3} M at 37°C) reflect a strong temperature dependence for dimerization. The K_d values were used to calculate the fraction monomer at the protein concentration used in the CD experiments and plotted at the respective temperatures as filled circles. The two methods of determining fraction monomer gave similar results.

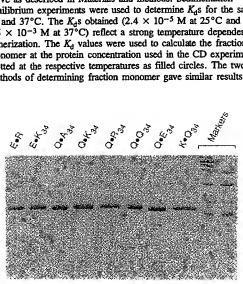

Fig. 5. Protein gel of eight purified proteins. Each protein is identified above the respective lane. We expressed and purified the mutant bZIP proteins from *E.coli* to >98% purity. Equal amounts of each protein, as determined spectroscopically, were loaded on a 14% SDS–polyacrylamide gel (Laemmli, 1970) and stained with Coomassie blue.

melting and a two-state transition upon thermal melting, thermodynamic parameters could be calculated (see Materials and methods).

The helicity, as measured by CD, was assumed to represent a two-state monomer–dimer equilibrium. The fraction monomer at different temperatures calculated from CD data was compared to the fraction monomer directly measured in the ultracentrifuge and found to be coincident (Figure 4). Thus, both CD measurements of helicity and sedimentation equilibrium measurements of molecular mass appear to monitor the same physical phenomenon: the melting of helical dimers into random coil monomers.

CD thermal denaturation was performed at different protein concentrations. The lowering of the melting temperature (T_m) at decreasing protein concentrations is consistent with the hypothesized monomer–dimer equilibrium (see Table I).

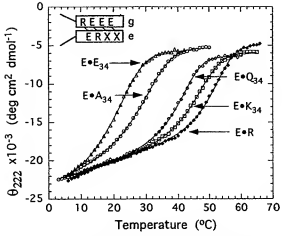

Fig. 6. CD thermal melting curves for the $E{\cdot}X_{34}$ proteins where X = A, R, K, Q or E. The line through each of the five labeled curves is a fitted curve as described in Materials and methods. Note that the initial baseline for each of the five curves is similar, indicating that all the samples start with the same fraction of dimer, which was shown to be 100% by sedimentation equilibrium. The cartoon in the upper left part of the figure is a graphic presentation of the g and e′ amino acid identities, and the presumed interactions are indicated by a line connecting the amino acids.

Thermodynamic stability of different g⟷e′ pairs

We generated a large collection of proteins mutated in the g and e′ positions. The energetics of the structurally observed g⟷e′ interactions were calculated based on the fact that our host protein exhibited a simple two-state transition upon thermal denaturation. The aim was to generate protein design rules based on a thermodynamic scale for various g⟷e′ amino acid pairs. Examination of the e and g positions of bZIP proteins shows that >80% of these positions are occupied by only four amino acids (Vinson *et al.*, 1993): K, R, E and Q. These represent three of the four charged amino acids and, perhaps more importantly, they contain long hydrophobic side chains. This may be essential for hydrophobic packing over the hydrophobic core created by the a and d positions (O'Shea *et al.*, 1991).

Six amino acids in the g and following e positions were systematically varied, namely the four commonly occurring amino acids in these positions, K, R, Q and E, as well as the only other remaining charged amino acid D, and finally A, a truncated amino acid used as a reference. Amino acids were changed in the fourth pair (two proteins), the third and fourth pairs (25 proteins), and the first, second, third and fourth pairs (six proteins). Figure 5 shows a representative protein gel of a number of the proteins purified from *E.coli*. All proteins were similarly pure.

The raw CD data for the thermal melting of five different $E{\cdot}X_{34}$ proteins are seen in Figure 6. In these samples mutated in the third and fourth g⟷e′ pairs, the g position is always occupied by E, while the e position is occupied by X = A, R, K, Q or E. All five samples have the same molar ellipticity at low temperatures, suggesting that each is totally dimeric at the beginning of the thermal melting, an assertion supported by ultracentrifugation data. At low temperatures, all five samples show a similar initial CD linear baseline followed by a clear co-operative transition indicative of melting of the helical leucine zipper. After thermal denaturation, all samples again have a similar

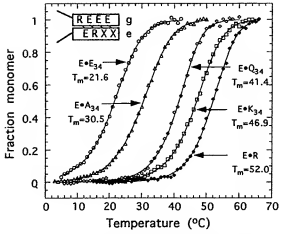

Fig. 7. The data from Figure 6 replotted as fraction monomer. The line through each curve is a fitted curve. The T_m for each of the five curves is noted.

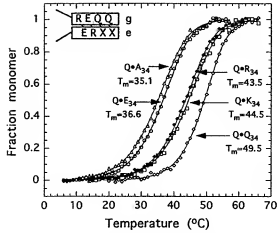

Fig. 8. CD thermal melts for Q·X_{34} proteins where X = A, R, K, Q or E, plotted as fraction monomer. The line through each curve is a fitted curve. The T_m for each of the five curves is noted. Note that the curves are more clustered and shifted toward the higher temperatures relative to Figure 7.

ellipticity. The same data plotted in terms of fraction monomer are shown in Figure 7. There is a 30°C difference in melting temperature (T_m) between the most (E·R) and least (E·E_{34}) stable sample. E·R is more stable than E·K_{34}. Both E·R and E·K_{34} have similar interacting charges, but their relative stability is as different as that between E·K_{34} and E·Q_{34} which have no charge–charge attraction.

CD data were obtained for the thermal melting of five different Q·X_{34} proteins mutated in the third and fourth heptads where the g position was always occupied by Q, but the e position was again occupied by the same amino acids used previously, X = A, R, K, Q or E (Figure 8). There is only a 15°C difference in the T_ms between the most (Q·Q_{34}) and least (Q·A_{34}) stable samples. Q interacts more stably with the basic amino acids, K and R, than with the acidic amino acid E.

Table I presents thermodynamic parameters calculated from CD thermal melts for 33 proteins that have been systematically varied to reveal the energetics of different g⟷e′ pairs. Each of the samples examined have different melting temperatures, indicating that changing amino acids in the g and e positions affects protein stability. Most samples retained their dimeric properties. The free energy differences due to mutations of E⟷R pairs to other g⟷e′ pairs were calculated from the difference in stability between E·R and mutant proteins. Mixing of two different proteins can create heterodimers with novel g⟷e′ pairs, which have been included in Table I. For example, the E·E_{34}/R·R_{34} heterodimer contains both E⟷R and R⟷E pairs. The difference between the E⟷R homodimer and E·E_{34}/R·R_{34} heterodimer is the replacement of two E⟷R with two R⟷E pairs. Thus, the difference in stability between these two dimers is the energetic difference between the E⟷R and R⟷E pair. This allows the calculation of the energy contribution of a R⟷E pair to stability.

The generality of the stability of g⟷e′ pairs ($\Delta\Delta G$) calculated from molecules mutated in the third and fourth heptads was addressed by creating proteins with changes in either only the fourth pair or all four pairs (Figure 9). The results show that, except for Q⟷Q, the calculated $\Delta\Delta G$ values per g⟷e′ pair are nearly independent of the heptad within which they reside.

Stability of g⟷e′ pairs relative to A⟷A

Table II presents a matrix of g⟷e′ pair stability for all combinations of amino acids investigated. The $\Delta\Delta G_{E\cdot R}$ values, found in Table I, for proteins containing altered third and fourth g⟷e′ pairs were recalculated relative to A⟷A. The A⟷A pair was chosen as the baseline because A is a truncated amino acid. A is unable to interact with its partner in a g⟷e′ pair since the β-carbons of the g⟷e′ pairs are 9 Å apart, in fact the presence of a single A precludes g⟷e′ interactions (O'Shea *et al.*, 1991). The most appropriate control protein, A·A_{34}, was insoluble. A·A_4, however, is soluble and has an interpretable CD thermal denaturation profile [A·A_3 was insoluble in low salt, but not high (1 M KCl) salt]. Since thermodynamic parameters were calculated per g⟷e′ pair, A·A_4 could be used to determine a reference point for $\Delta\Delta G$.

Because of the importance of the energetic value of the A⟷A pair in subsequent calculations of coupling energies between amino acids in the g⟷e′ pair, we independently calculated the value of the A⟷A pair using two mixing experiments. A mixture of E·A_{34} and A·R_{34} formed a heterodimer, as evidenced by the one-stage melting transition with a T_m higher than either homodimer. The structural difference between this heterodimer and the E·R homodimer is the replacement of an E⟷R pair in both the third and fourth heptad with A⟷A pairs. The difference in ΔG_o between an E·A_{34}/A·R_{34} heterodimer and E·R homodimer of 2.82 kcal/mol is a measure of the strength of an A⟷A pair relative to an E⟷R pair in both the third and fourth heptads. The average value of 1.41 kcal/mol/pair is similar to the value of 1.33 calculated for the A⟷A pair from A·A_4. A similar type of experiment examined the difference in ΔG_o between the heterodimer A·E_{34}/R·A_{34} and E·B_{34}/R·R_{34} which again differ by an A⟷A pair in both the third and fourth heptad. This analysis suggests that A⟷A is 1.46 kcal/mol less stable than E⟷R. The similarity between the values obtained for the A⟷A pair by melting A·A_4 and the two heterodimers containing A⟷A pairs gave us a greater confidence in the calculation of the energetic strength of the A⟷A pair.

Table I. Fitted thermodynamic parameters calculated for the thermal unfolding of different VBP g⟷e′ derivatives per mole of dimer

Sample 3.4 μM	T_m (°C)	$\Delta H(T_m)$ (kcal/mol)	$\Delta G(37)$ (kcal/mol)	$\Delta\Delta G_{E-R}$ (kcal/mol)	$K_d(37)$ (mol)	Mw[a] (kDa)
E·R	52.0 ± 0.1	−83	−11.2	0.00	1.4e-08	21.9
EK_4	49.7 ± 0.1	−74	−10.4	0.40	5.2e-08	
$Q{\cdot}Q_{34}$	49.3 ± 0.2	−80	−10.6	0.16	3.9e-08	21.7
$E{\cdot}K_{34}$	47.0 ± 0.1	−70	−9.80	0.35	1.4e-07	
$Q{\cdot}K_{34}$	44.5 ± 0.2	−63	−9.18	0.50	3.7e-07	21.8
$Q{\cdot}R_{34}$	43.5 ± 0.1	−64	−9.04	0.54	4.7e-07	22.5
$E{\cdot}K_{1234}$	42.3 ± 0.2	−67	−8.87	0.29	6.1e-07	21.6
$A{\cdot}Q_{34}$	42.0 ± 0.1	−71	−8.88	0.58	6.0e-07	
$K{\cdot}Q_{34}$	41.9 ± 0.2	−61	−8.70	0.62	8.0e-07	22.1
$E{\cdot}Q_{34}$	41.3 ± 0.1	−77	−8.82	0.60	6.6e-07	18.8
$A{\cdot}K_{34}$	41.3 ± 0.2	−68	−8.70	0.63	8.1e-07	21.1
$A{\cdot}A_4$	40.3 ± 0.2	−73	−8.55	1.33	1.0e-06	21.6
$R{\cdot}Q_{34}$	38.9 ± 0.2	−65	−8.19	0.75	1.8e-06	
$Q{\cdot}E_{34}$	36.6 ± 0.2	−63	−7.72	0.87	3.9e-06	22.6
$Q{\cdot}A_{34}$	35.1 ± 0.3	−58	−7.44	0.94	6.2e-06	20.2
$Q{\cdot}R_{1234}$	35.0 ± 0.2	−64	−7.38	0.48	6.8e-06	
$Q{\cdot}K_{1234}$	34.0 ± 0.2	−62	−7.18	0.50	9.4e-06	
$K{\cdot}R_{34}$	33.4 ± 0.8	−45	−7.25	0.99	8.4e-06	
$A{\cdot}E_{34}$	33.0 ± 0.2	−63	−6.95	1.1	1.4e-05	17.3
$K{\cdot}R_{34}$	33.0 ± 0.2	−47	−7.16	1.0	9.7e-06	
$Q{\cdot}Q_{1234}$	32.8 ± 0.2	−56	−7.00	0.53	1.3e-05	18.9
$K{\cdot}A_{34}$	31.5 ± 0.2	−51	−6.84	1.09	1.6e-05	
$R{\cdot}A_{34}$	30.7 ± 0.2	−49	−6.71	1.12	2.0e-05	20.9
$E{\cdot}A_{34}$	30.5 ± 0.2	−60	−6.43	1.19	3.1e-05	21.2
$R{\cdot}R_{34}$	27.1 ± 0.4	−40	−6.29	1.23	3.9e-05	
$R{\cdot}K_{34}$	26.2 ± 0.8	−37	−6.24	1.24	4.2e-05	
$A{\cdot}R_{34}$	25.3 ± 0.8	−37	−6.09	1.28	5.5e-05	22.5
$E{\cdot}E_{34}$	21.6 ± 0.4	−56	−4.41	1.70	8.1e-04	21.1
$E{\cdot}Q_{1234}$	20.6 ± 0.5	−50	−4.49	0.84	7.2e-04	
$D{\cdot}R_{34}$	17.5 ± 1.5	−40	−4.37	1.71	8.7e-04	
$D{\cdot}A_{34}$	<0.00	(−40)[b]	0.41	>2.90	1.9e + 00	
$D{\cdot}D_{34}$	<0.00					
$D{\cdot}K_{34}$	15.5 ± 0.5	−47	−3.39	1.95	4.2e-03	
$Q{\cdot}A_{1234}$	21.7 ± 1.5					36.7[c] (26.3)
E·R C-S	40.6 ± 0.2	−71	−8.59		9.6e-07	21.8
$E{\cdot}K_{1234}$ C-S	30.1 ± 0.2	−44	−6.71		2.0e-05	21.3
$E{\cdot}K_4$ (1 μM)	44.9 ± 0.3	−74	−10.3	0.46	6.3e-08	
$E{\cdot}K_4$ (8.2 μM)	51.6 ± 0.1	−78	−10.4	0.41	5.4e-08	
$E{\cdot}K_4$ (27.5 μM)	54.6 ± 0.1	−82	−10.4	0.43	5.6e-08	
$E{\cdot}E_{34}/RK_{34}$	50.3 ± 0.1	−80	−11.2		1.4e-08	
$A{\cdot}E_{34}/RA_{34}$	41.5 ± 0.2	−68	−8.74		7.6e-07	21.6
$E{\cdot}A_{34}/AR_{34}$	39.9 ± 0.2	−63	−8.37		1.4e-06	
$E{\cdot}E_{34}/KR_{34}$	50.6 ± 0.1	−83	−11.4		1.1e-08	
$E{\cdot}E_{34}/RR_{34}$	51.5 ± 0.1	−85	−11.6		7.0e-09	

The table presents energetic calculations for a variety of mutant VBP proteins. The following parameters are presented: melting temperature, T_m and curve fitting error; dimerization free energy extrapolated to T = 37°C, $\Delta G(37)$; dimerization van't Hoff enthalpy at $T = T_m$, $\Delta H(T_m)$; dissociation constant at 37°C, $K_d(37)$; the mol. wt as determined by sedimentation equilibrium, Mw; the energetic difference in a single g⟷e′ pair relative to E⟷R, $\Delta\Delta G_{E-R}$. Samples are arranged in order of decreasing thermal stability of the non-covalently linked protein. All the samples reported here have a similar ellipticity at low temperatures, suggesting that they are dimeric before thermal denaturation. The standard errors for ΔG_s vary from ±0.04 to ±0.16 kcal/mol, except for $D{\cdot}R_{34}$ (0.38 kcal/mol).

[a]The mol. wt (Mw) of the E·R monomer is 11 256 Da.

[b]ΔH was assumed to allow a calculation of ΔG.

[c]This sample fits a monomer⟷tetramer equilibrium equation. The value in parentheses is for the oxidized sample.

Coupling energies ($\Delta\Delta G_{int}$) of g⟷e′ pairs

The data described so far do not demonstrate that the calculated differences in energy are necessarily caused by the interaction of amino acids in the e and g positions. These differences could simply be the sum of the independent energetic contributions of each amino acid in the g⟷e′ pair. Evidence for an interaction between amino acids in g⟷e′ pairs was obtained from two arguments. The first is based on a thermodynamic cycle and the second is from mixing experiments discussed in the next section. A thermodynamic cycle required the generation of proteins with non-interacting g⟷e′ pairs. We generated a set of proteins with A in either the e, g, or e and g positions, e.g. $E{\cdot}A_{34}$, $A{\cdot}R_{34}$ and $A{\cdot}A_4$. By comparing the stability of these proteins with proteins in which the e and g positions are occupied by potentially interacting amino acids, e.g. E·R, we were able to calculate an energy of interaction between E and R, termed the coupling energy ($\Delta\Delta G_{int}$).

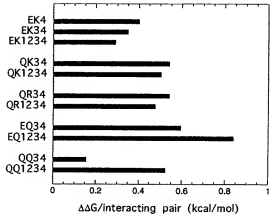

Fig. 9. Histogram of the calculated energetic strength of different g⟶e′ pairs relative to E⟶R ($\Delta\Delta G_{E\cdot R}$) for proteins with different numbers of a particular g⟶e′ pair. The top of the figure shows three proteins containing one, two or four E⟶K pairs. The average difference in stability is consistently around 0.4 kcal/mol relative to E⟶R, suggesting that the measured difference is independent of the location of the particular g⟶e′ pair along the length of the leucine zipper. Note that in the constructs that contain an altered first heptad, the order of the amino acid pair is reversed, as is true for the host protein. The general conclusion is that the calculated stability of a particular g⟶e′ pair is similar for proteins in which all four heptads have the same amino acid pair or only the third and fourth heptads have the same pair. The one notable exception is Q⟶Q; $Q\cdot Q_{34}$ is very stable, while $Q\cdot Q_{1234}$ is much less stable than expected, suggesting that some particular structural interaction is occurring in one context, but not the other.

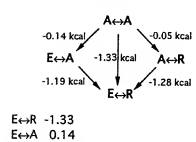

Fig. 10. Thermodynamic cycle for the interaction of glutamic acid in the g position with arginine in the following e′ position. The four proteins in the cycle are $A\cdot A_4$, $E\cdot A_{34}$, $A\cdot R_{34}$ and $E\cdot R$. The $\Delta\Delta G$ values presented are in terms of an individual g⟶e′ interaction. The E⟶R pair is 1.33 kcal/mol more stable than A⟶A. The contribution of the individual amino acids to the stability of the leucine zipper was determined by studying $E\cdot A_{34}$ and $A\cdot R_{34}$. The E⟶A pair is 0.14 kcal/mol more stable than A⟶A. The A⟶R pair is 0.05 kcal/mol more stable than A⟶A. The sum of the individual contributions of E and R to the dimer stability is −0.19 kcal/mol. The extra −1.14 kcal/mol of stability [−1.33−(−0.19)] from the E⟶R pair is the coupling energy (ΔG_{int}) indicative of the interaction of E with R across the leucine zipper.

Table II. Thermodynamic differences for g⟶e′ interactions relative to A⟶A ($\Delta\Delta G_{A\cdot A}$) (kcal/mol)

i g/e′⟶	A	E	Q	R	K
A	0.0	−0.27	−0.75	−0.05	−0.70
E	−0.14	+0.37	−0.73	−1.33	−0.98
Q	−0.39	−0.46	−1.17	−0.79	−0.83
R	−0.21	−1.55[a]	−0.58	−0.10	−0.10
K	−0.24	−1.42[a]	−0.71	−0.32	−0.34
D	>+1.57	–	–	+0.38	+0.62

The $\Delta\Delta G_{E\cdot R}$ values from Table I were recalculated relative to the A⟶A pair to produce $\Delta\Delta G_{A\cdot A}$. The standard errors for $\Delta\Delta G_0$ vary from ±0.01 to ±0.04 kcal/mol except for $D\cdot R_{34}$ (0.1 kcal/mol).
[a]The R⟶E and K⟶E values were calculated from the melting of heterodimers (see Table I).

Table III. Coupling energy ($\Delta\Delta G_{int}$) of g⟶e′ pair (kcal/mol)

i g/e′⟶	E	Q	R	K
E	+0.78	+0.16	−1.14	−0.14
Q	+0.20	−0.03	−0.35	+0.26
R	−1.07[b]	+0.38	+0.16	+0.81
K	−0.91[a]	+0.28	−0.03	+0.60

The coupling energy ($\Delta\Delta G_{int}$) was calculated from values given in Table II. The $\Delta\Delta G$ values for both the X⟶A and A⟶Y pairs were subtracted from the value for the X⟶Y pair to determine the X⟶Y coupling energy ($\Delta\Delta G_{int}$).
[a]as Table II.

A thermodynamic cycle for the E⟷R pair is presented in Figure 10. As noted earlier, the E⟷R pair is −1.33 kcal/mol more stable than A⟷A. To determine the coupling energy ($\Delta\Delta G_{int}$) between E and R, the independent contributions of either E or R alone to the stability of the leucine zipper had to be determined. This was accomplished by examining the stability of $E\cdot A_{34}$ and $A\cdot R_{34}$. The difference in stability between an A⟷A pair and other A-containing pairs (e.g. A⟷R) shows the ability of the non-alanine amino acid to stabilize the dimer independently of a g⟷e′ interaction. The E⟷A pair is −0.14 kcal/mol more stable than A⟷A and the A⟷R pair is −0.05 kcal/mol more stable than A⟷A. Subtracting these values from the measured −1.33 kcal/mol for the E⟷R pair gives a value of −1.14 kcal/mol for the coupling energy ($\Delta\Delta G_{int}$) of E interacting with R.

Coupling energies for 16 g⟷e′ pairs are presented in Table III. In contrast to the E⟷R pair, the E⟷K pair shows a negligible coupling energy of −0.14 kcal/mol. A general trend from these coupling energy calculations is that g⟷e′ pairs containing R in the e position have stronger (more negative) coupling energies than those containing K in the e position. An observation that was not expected from model building is the large number of positive coupling energies. These suggest that there is some sort of repulsion or steric clash between amino acids trying to occupy the same hydrophobic patch over the leucine zipper core. The order of amino acids in the g⟷e′ pair also affects the coupling energy. For example, K⟷R shows a negligible coupling energy of −0.03 kcal/mol, while R⟷K shows the strongest measured repulsive coupling energy of +0.81 kcal/mol. This difference suggests that the order of amino acids in a g⟷e′ pair is critical for the amino acid interactions. The order-dependent coupling energy is not seen for all g⟷e′ pairs, the K⟷Q and Q⟷K pairs have a similar coupling energy of +0.27 kcal/mol.

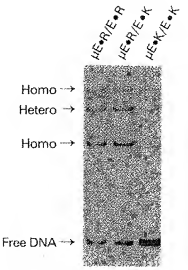

Fig. 11. Gel-retardation assay mixing two different size proteins. Equal moles of the four proteins named in the figure were mixed as indicated. Note that the mixing of $\mu E{\cdot}R$ with $E{\cdot}K_{1234}$ results in a similar proportion of heterodimer formed as in either the $\mu E{\cdot}R$ with $E{\cdot}R$ or $\mu E{\cdot}K_{1234}$ with $E{\cdot}K_{1234}$. $\mu E{\cdot}R$ binds DNA about twice as well as $E{\cdot}R$ which explains the greater DNA shifting of $\mu E{\cdot}R$ relative to $E{\cdot}R$. Also, $E{\cdot}K_{1234}$ binds DNA less well than $E{\cdot}R$.

Novel heterologous interactions regulate dimerization specificity

Amino acids in the g and e positions have been implicated in regulating dimerization specificity. Since the mutant proteins described here involve changes only in these positions, they were used in mixing experiments to investigate the specificity of hetero versus homodimerization. If we mixed $E{\cdot}R$ and $E{\cdot}K_{1234}$, the heterodimer would contain an equal number of E⟷R and E⟷K interactions which are the same type of interactions present in the respective homodimers. A thermodynamic analysis, assuming g⟷e′ pairs are independent of each other, reveals that the stability of the heterodimer would be the average of the stabilities of the two homodimers. This suggestion was tested using a gel-shift assay. A short version of the protein $E{\cdot}R$ was independently mixed with two proteins, a large $E{\cdot}R$ and a large $E{\cdot}K_{1234}$, bound to DNA and resolved on a native polyacrylamide gel (Figure 11). The formation of heterodimer is similar in all three lanes. The $E{\cdot}R/E{\cdot}K_{1234}$ heterodimer contains an equal number of E⟷R and E⟷K interactions, and consequently would be expected to be intermediate in stability between $E{\cdot}R$ and $E{\cdot}K_{1234}$, 1.2 kcal/mol less stable than $E{\cdot}R$. One might have expected that mixing of $E{\cdot}R$ with a protein 2.3 kcal/mol less stable ($E{\cdot}K_{1234}$) would result in less formation of heterodimers. This is not the result; the laws of mass action allow a less stable protein ($E{\cdot}K_{1234}$) to disrupt the stability of a more stable protein ($E{\cdot}R$) if the heterodimer is of intermediate stability.

In the second mixing experiment, the stability of the heterodimer is calculated to be greater than the average of the two homodimer stabilities, thus favoring the formation of heterodimers. Mixing E⟷E and R⟷R would result in a heterodimer containing E⟷R and R⟷E interactions which are not present in the homodimers. Two proteins

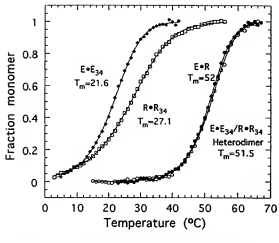

Fig. 12. Mixing of $E{\cdot}E_{34}$ with $R{\cdot}R_{34}$. This plot shows the CD thermal melting of four samples as labeled. They are $E{\cdot}E_{34}$, $R{\cdot}R_{34}$, $E{\cdot}R$, and an equimolar mixture of $E{\cdot}E_{34}$ with $R{\cdot}R_{34}$. The equimolar mixture of $E{\cdot}E_{34}$ and $R{\cdot}R_{34}$ melts as a single stable species, much more stable than either alone, suggesting that these proteins form heterodimers. The presumed $E{\cdot}E_{34}/R{\cdot}R_{34}$ heterodimer has a melting profile similar to $E{\cdot}R$. The total protein concentration for all samples was 3.4 μM.

($E{\cdot}E_{34}$ and $R{\cdot}R_{34}$) were mixed and shown to be more thermally stable than solutions of either protein alone (Figure 12) (Graddis *et al.*, 1993; O'Shea *et al.*, 1993). In fact, the mixture of $E{\cdot}E_{34}$ and $R{\cdot}R_{34}$ has similar stability to $E{\cdot}R$. We argue that the increased stability comes from the novel pair of amino acids in the g and e position present only in the heterodimer. The interaction could be either attractive between dissimilar monomers, encouraging heterodimers, or due to repulsion between similar monomers discouraging homodimers, or a combination of both.

A in the e position can create tetramers

Eighteen mutant proteins were analyzed in the analytical ultracentrifuge, at temperatures below the transition temperature, to determine their oligomerization properties. Most are dimers with surprisingly small variations in the calculated mol. wts. $Q{\cdot}A_{1234}$, however, did show a dramatic change in mol. wt. Analysis of sedimentation data revealed that this molecule behaves as a tetramer (Figure 13). However, $Q{\cdot}A_{34}$, a similar protein to $Q{\cdot}A_{1234}$, behaves as a dimer. It is likely that the conversion to tetramer is caused by the large hydrophobic face present in $Q{\cdot}A_{1234}$. Oxidized $Q{\cdot}A_{1234}$, corresponding to a covalently linked leucine zipper, is dimeric, while the reduced $Q{\cdot}A_{1234}$ sample is tetrameric. This indicates that the tetramer is not two interacting leucine zipper dimers. Similarly, the e and g positions of the GCN4 leucine zipper have been mutated to A, attached to a DNA binding domain and found to form tetramers (Alberti *et al.*, 1993).

Discussion

This paper reports the thermal stability of 33 leucine zipper proteins containing 27 different systematic combinations of amino acids in the g position on one helix and the following e position on the opposite helix (denoted g⟷e′). CD thermal stability measurements have allowed us to determine the

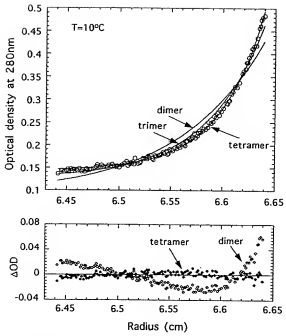

Fig. 13. Sedimentation equilibrium determination of the mol. wt of $Q{\cdot}A_{1234}$ at 10°C. The sample was in 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 1 mM EDTA, 1 mM DTT. Each sample was loaded at three concentrations (10, 20 and 40 μM) which have ODs of 0.1, 0.2 and 0.4 at 280 nm. Samples were spun at 25 000 r.p.m. for 24 h. Theoretical curves for a dimer, trimer and tetramer are plotted as solid lines. The actual data are plotted as circles. These data clearly fit onto the tetramer curve. The bottom panel shows the residual plots of fitting the experimental data to a dimer model or tetramer model. The tetramer model shows less systematic error in the residual plot, indicating that it fits the data better. If the sample was prepared and spun in the absence of DTT, the data obtained fitted a dimer curve.

energetic contribution of a single g⟷e′ amino acid pair and, based on a thermodynamic cycle analysis, formally determine the coupling energy ($\Delta\Delta G_{int}$) between the amino acids in g⟷e′ pairs. The results have provided some general rules concerning g⟷e′ pairs. E⟷R is the most stable g⟷e′ pair. R has a wider range of dimerization stabilities than K, suggesting more potential for specificity in leucine zipper dimerization. E has a wider range of dimerization stabilities than Q, which generally forms stable g⟷e′ pairs even though it is uncharged. These rules should allow the design of leucine zipper proteins that would specifically dimerize with target leucine zipper helices, but not with others.

ΔΔG of g⟷e′ pairs is position independent

An important concern in these experiments is whether or not the g⟷e′ pairs studied were energetically independent of the rest of the leucine zipper structure. It could be argued that each heptad is structurally unique, and only a well-defined heptad should be studied. We took an alternative approach of examining e and g pairs in different structural environments, and taking an average in the hope of producing more general design rules. Additionally, we wanted to amplify any change in stability by changing g⟷e′ pairs in two different heptads, which represents four interactions: two interactions per heptad, one on each side of the leucine zipper. Examination of the third and fourth heptads of VBP (Figure 1) shows that besides the charged amino acids in the g and e positions, charged amino acids are also present in the c position. The third heptad has an E and the fourth heptad has an R. Work by several groups has shown that intrahelical salt bridges, both $(i,i+3)$ and $(i,i+4)$, can exist. Thus, the possibility exists of a $(i,i+3)$ type intrahelical salt bridge existing between the g and following c position in both the third and fourth heptad (Marqusee and Baldwin, 1987; Fairman *et al.*, 1990; Gans *et al.*, 1991; Scholtz *et al.*, 1993). Fortunately, the c amino acids in the third and fourth heptads are of opposite charge, hopefully canceling out any systematic error in these experiments caused by the c amino acids having the same charge. The crystal structure of the leucine zipper protein, GCN4, shows that E(270) in the g position appears to form a g⟷e′ interhelical interaction with K(275), and an $(i,i+3)$ intrahelical interaction with R(273). This observation suggests that the thermal stability being measured may be a complex combination of interactions between the g and the following c and e′ positions.

To determine if the g⟷e′ pairs examined were acting independently of the local leucine zipper structure, we examined proteins containing different numbers of a particular g⟷e′ pair and determined the average strength of each pair. In most cases, similar $\Delta\Delta G$ values per g⟷e′ pair were obtained for each of five different g⟷e′ pairs present in one, two or four repeats. Since the g and e amino acids pack over the hydrophobic core, particularly the a position (O'Shea *et al.*, 1991), the identity of the amino acid in the a position may influence the energetics of the g⟷e′ interaction. The results presented in Figure 9 support the general conclusion that the $\Delta\Delta G$ values per g⟷e′ pair are independent of position along the leucine zipper, even though the first and second heptads have an alanine and asparagine, respectively, in the a position, while the third and fourth heptads have a valine. This suggests that the contribution of the a position to g⟷e′ pair stability is not dramatic. The only exception to this result is Q⟷Q. The $Q{\cdot}Q_{34}$ protein indicates that Q⟷Q is much stronger than anticipated from examination of $Q{\cdot}Q_{1234}$. We do not know if these results over- or under-represent the actual Q⟷Q interaction energy. No simple interpretation of this result is obvious.

Stabilizing contributions of amino acids in the e or g positions

The contribution of individual amino acids in the e and g positions to the formation of leucine zippers can be conceptually divided into three distinct effects: oligomerization state, inherent dimer propensity and dimerization specificity. The first effect is the propensity to form different oligomeric states. Muller-Hill's group has shown that changing amino acids in the e and g positions switches the oligomerization between dimers and tetramers (Alberti *et al.*, 1993). Our own work with the protein $Q{\cdot}A_{1234}$ has shown that placing an A in each of the four e positions can create tetramers.

The second contribution of amino acids in the e and g positions is the inherent propensity of a particular amino acid to form a dimeric coiled coil. This is analogous to the α-helix-forming propensities of amino acids for isolated α-helices (Lye *et al.*, 1990; Padmanabhan *et al.*, 1990; Blaber

et al., 1993) and solvent-exposed positions (the **f** position) of a coiled coil (O'Neil and DeGrado, 1990). The stability of A↔X- or X↔A-containing proteins allows the determination of the energetic contribution of the X amino acid in either the **g** or e position in presumed isolation from any **g**↔e′ interaction. These stabilities could reflect the α-helical propensity of a particular amino acid. The difference between the α-helical propensity of a particular amino acid and our data for the e and **g** positions reflects the structural consequence of these two positions being near the hydrophobic interface of the leucine zipper. For example, A↔K is 0.70 kcal/mol more stable than A↔A even though K forms helices 0.12 kcal/mol more poorly than A (O'Neil and DeGrado, 1990). This difference of 0.82 kcal/mol reflects the ability of K in the **e** position to stabilize the leucine zipper independent of α-helix propensity. This is probably due to the methylenes of the lysine side chain packing over the hydrophobic core of the leucine zipper. Similar, though less dramatic, results are seen with the amino acids E, Q and R, all of which have long side chains. However, aspartic acid forms helices 0.62 kcal/mol more poorly than A, but in the **g** position is >1.5 kcal/mol less stabilizing than A. This additional destabilization of D in the **g** position suggests that the short side chain of D may cause destabilization greater than expected based on α-helical propensity.

Calculating the strength of a **g**↔e′ pair relative to A↔A allows an understanding of the contribution of the **g**↔e′ pair to dimer stability. E↔R is 1.33 kcal/mol more stable than A↔A. Thus, the protein E·R containing eight **g**↔e′ pairs over four heptads would be 10.6 kcal/mol more stable than $A \cdot A_{1234}$. One general observation is that most of the **g**↔e′ pairs we have examined are more stable than the A↔A pair. This even extends to some of the homologous interactions that might be expected to be repulsive. For example, K↔K is 0.34 kcal/mol more stable than A↔A. Presumably, repulsive effects between the two positively charged lysine amino acids exist, but are compensated for by the hydrophobic packing of the long K side chain across the dimerization interface. Similar results are seen for the R↔R pair. Only E↔E has a repulsive interaction and this pair is 0.37 kcal/mol less stable than A↔A.

The third effect of particular amino acids in the **e** and **g** positions is to regulate dimerization specificity, a direct result of interhelical interactions between the **g** and e′ amino acids, discussed in detail below.

Interactions between g and e′ amino acids

The use of a thermodynamic cycle analysis allowed us to unravel the interaction energy between the two amino acids in the **g**↔e′ pair; this is termed the coupling energy ($\Delta\Delta G_{int}$) (Carter *et al.*, 1984; Horovitz *et al.*, 1990; Serrano *et al.*, 1990). This analysis led to perhaps the most interesting result of this work (Table III). A simple comparison of the E↔R and E↔K pairs indicates that E↔R is 0.35 kcal/mol more stable than E↔K, the same trend is seen for intrahelical E↔R and E↔K pairs (Merutka and Stellwagen, 1991). An examination of the stabilizing contribution of K and R alone (A↔K and A↔R) indicates that K is 0.65 kcal/mol more stabilizing than R. The calculation of a coupling energy for each **g**↔e′ pair indicates that E↔R has a $\Delta\Delta G_{int}$ of 1.14 kcal/mol, while E↔K has a $\Delta\Delta G_{int}$ of only 0.14 kcal/mol. The same general trend is seen when comparing the coupling energies of E and Q with K and R. In each case, R-containing pairs have a stronger $\Delta\Delta G_{int}$ than K-containing pairs. The positive coupling energy of the Q↔K pair may represent a steric clash between the two amino acids as they hydrophobically pack over the **a** and **d** positions.

The very low coupling energy of the E↔K pair (−0.14 kcal/mol) is consistent with data from Kim's group (O'Shea *et al.*, 1992). These investigators examined the structural specificity required for the preferential heterodimerization of the two bZIP proteins Fos and Jun. They were able to map part of the heterodimerization specificity to the **e** and **g** position of each protein; Fos contains an E↔E pair, while Jun contains a K↔K pair and the heterodimer contains both an E↔K and a K↔E pair. Lowering the pH to 2 resulted in no change in the dimerization strength of the heterodimer, suggesting the absence of any ion pair attraction in the heterodimer. The absence of any coupling energy between the E↔K pair and some coupling energy in the calculated K↔E pair is consistent with the pH experiment.

Structural rational for coupling energy ($\Delta\Delta G_{int}$)

The physical forces that result in the measured coupling energies are unclear. Salt and pH experiments in progress should help to address this question. The type of forces that could exist are van der Waals packing between amino acids, hydrogen bonding between charged or uncharged amino acids (salt independent) or charge−charge interactions (salt dependent) (Scholtz *et al.*, 1993).

Table IV presents a comparison between the stability of E↔X and Q↔X pairs (see Figures 7 and 8). Q is structurally similar to a protonated E, which can be created at low pH; both can accept and donate a hydrogen bond. Two structural differences exist between Q and a charged E. Q accepts and donates a hydrogen bond, while E accepts two hydrogen bonds in addition to having possible charge−charge interactions. The difference between the $\Delta\Delta G$ for E↔X and Q↔X pairs highlights how these two amino acids interact with the basic amino acids K and R. E↔K is only −0.15 kcal/mol more stable than Q↔K while is −0.54 kcal/mol more stable than Q↔R. E displays a greater range of interaction energies than Q primarily because E↔E is a very destabilizing interaction. The small energetic difference between E↔K and Q↔K (−0.15 kcal/mol) suggests that the possible charge−charge attraction between E and K is not a major contributor to the stability of the pair. The near absence of any coupling energy between E and K ($\Delta\Delta G_{int}$ = −0.14 kcal/mol) independently confirms the absence of any charge−charge interactions. The

Table IV. Energetic difference between E—X and Q—X pairs (kcal/mol)

i**g**/e′→	A	E	Q	R	K
E	−0.14	+0.37	−0.73	−1.33	−0.98
Q	−0.39	−0.46	−1.17	−0.79	−0.83
E−Q	+0.25	+0.83	+0.44	−0.54	−0.15

ΔG values were calculated from CD thermal melts and normalized to A↔A. The differences between the energetics of E↔X and Q↔X pairs have been calculated (E−Q) to emphasize the difference exhibited by E relative to Q.

large energetic difference between E⟷R and Q⟷R (−0.54 kcal/mol) may reflect the charge−charge contribution to the stability of the E⟷R pair and/or the formation of a second hydrogen bond between the guanidinium group of R and E that is not available with Q. The coupling energy ($\Delta\Delta G_{98}$) between E and R of −1.14 kcal/mol is consistent with a charge interaction and two possible hydrogen bonds between E and R. Electrostatic interactions have been observed in a variety of proteins (Horovitz *et al.*, 1990; Dao-pin *et al.*, 1991; Schmidt-Dor *et al.*, 1991; Robinson and Sligar, 1993). The magnitude of the interactions observed by other investigators is similar to the results presented here. Although individual interaction energies are small, they are important to leucine zipper stability and dimerization specificity because of the large number of such interactions that are possible along the dimer interface.

Heterologous interactions are critical for dimerization specificity

The fact that the strongest **g⟷e′** pair (E⟷R) consists of different amino acids allows the design of proteins that preferentially homo or heterodimerize (Figure 12). The difference between the best (E⟷R) and the worst (E⟷E) **g⟷e′** pair is 1.7 kcal/mol. Changing a single R to E in each monomer would create a dimer with two new destabilizing interactions, one on each side of the leucine zipper, resulting in a dimer that is 3.4 kcal/mol less stable. This corresponds to a dimerization constant that is 240 times weaker. The energetic contribution of an E⟷R pair to leucine zipper stabilization (−2.66 kcal/mol/heptad) is similar to the contribution of leucine relative to alanine in the **a** (−3.3 kcal/mol) or **d** (−2.0 kcal/mol) position (Zhou *et al.*, 1992). Thus, the **g⟷e′** pair can contribute over one-third of the total stabilization of the leucine zipper. Additional experiments with heterologous pairs in the interior of the leucine zipper would be valuable to ascertain if heterologous interactions between **d⟷d′** or **a⟷a′** could help mediate dimerization specificity. It is expected that further experiments, examining more complex heptads, will provide further details on the structural independence of the **g⟷e′** pair.

The large number of leucine zipper proteins in mammalian systems suggests that a judicious combination of attractive and repulsive interactions may be needed to design specific dimerization partners. Since the specificity of dimerization is distributed throughout the length of the leucine zipper, the potential for modulation of dimerization partners is great. Our design of leucine zipper partners that preferentially heterodimerize can be a useful biological tool to bring together different cellular proteins at specific locations in the cell.

Materials and methods

Protein

The sequence of the 96 amino acid host protein is ASMTGGQQMGRDP-LEE-KVFVPDEQKD EKYWTRRKKN NVAAKRSRDA RRLKENQFTI RAAFLEKENT ALRTEVAELR KEVGRCKNIV SKYETRYGPL. The 'leucine' positions are in bold type. The first 13 amino acids are from φ10 (Studier and Moffatt, 1986), the next three amino acids are a cloning linker and the remaining 80 amino acids are the C-terminus of VBP (Iyer *et al.*, 1991), the chicken equivalent of the mammalian DBP gene (Mueller *et al.*, 1990). These 80 amino acids contain the entire bZIP region of the protein and are able to bind to DNA as a dimer in a sequence-specific manner (Vinson *et al.*, 1993). VBP contains a single cysteine in the d position near the C-terminus of the leucine zipper region. This cysteine can be oxidized to form more stable dimers than the reduced protein. Oxidized and reduced proteins display identical helicity at low temperatures. A disulfide bond between cysteines in the d position has previously been shown not to disrupt the helix-forming capacity of coiled coils (Zhou *et al.*, 1993). The covalently linked samples bind DNA with similar binding constants to the reduced samples, suggesting that they have similar conformations (data not shown). We have generated two proteins (E·K$_{1234}$ C-S and E·R C-S) where the cysteine has been replaced with serine and found that samples containing a reduced cysteine behave similarly to the serine-containing proteins.

Protein expression and purification

Proteins were synthesized in *E.coli* using the phage T7 expression system (Studier and Moffatt, 1986). Bacterial cultures (400 ml) at an OD of 0.6 at 600 nm were induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 2 h. Cells were recovered by centrifugation, resuspended in 6 ml of lysis buffer [50 mM Tris−HCl (pH 8.0), 1 mM EDTA, 1 mM benzamidine, 1 mM dithiothreitol (DTT) and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)], frozen, thawed and gently brought to 1 M KCl by the addition of 2 ml of 4 M KCl. The samples containing clusters of basic amino acids were gently brought to 2 M KCl before the initial centrifugation. The samples were centrifuged at 25 000 r.p.m. in a Beckman T42 rotor and the supernatant was isolated. The isolated supernatant was then heated to 65°C for 10 min, centrifuged and the supernatant again isolated. The proteins were dialyzed to 20 mM Tris−HCl (pH 8.0), 100 mM KCl, 1 mM EDTA and 1 mM DTT, and loaded onto a heparin agarose column. The column was washed with lysis buffer containing 100 mM KCl, followed by a 300 mM KCl wash, and eluted with buffer containing 1 M KCl. The samples were then dialyzed to 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 1 mM EDTA, with or without 1 mM DTT for 48 h with a change of buffer after 24 h. The purity of the proteins was assayed by SDS–PAGE (Laemmli, 1970). The different size proteins used in the gel-shift experiment (Figure 11) were described previously (Vinson *et al.*, 1993).

Construction of mutant proteins

Amino acid substitution mutants were introduced into DBP by the four-primer polymerase chain reaction (PCR) mutagenesis method (Ho *et al.*, 1989). DNA sequencing was performed on double-stranded templates using the Sanger dideoxynucleotide method (Sanger *et al.*, 1977).

Equilibrium sedimentation

Equilibrium sedimentation measurements were performed using a Beckman XL-A Optima Analytical Ultracentrifuge equipped with absorbance optics and a Beckman An-60Ti rotor. Samples were loaded at three concentrations, 10, 20 and 40 μM (0.1, 0.2 and 0.4 OD at 280 nm), into a six-hole centerpiece and spun at 25 000 r.p.m. for 24 h. Twenty data sets for three concentrations were averaged and jointly fitted for a singular mol. wt. Some calculations assumed a monomer−dimer equilibrium. Compositional partial specific volumes for the proteins were calculated according to Zamyatnin (1984). All scans were done at 25°C, except A·R$_{34}$ and A·K$_{34}$ which were done at 10°C.

Circular dichroism studies

CD studies were performed using a Jasco J-710 spectropolarimeter with a 5 mm rectangular CD cell. Temperature scans were performed by scanning continuously from 4 to 80°C at a scan rate of 1°C/min. Data were collected using the time scan mode of the Jasco-710 software. The ellipticity at 222 nm (θ_{222}) was recorded every 1 min with a response time of 5 s and a bandwidth of 1 nm. The temperature in the cell was controlled using a water-jacketed cell holder connected to a Haake F3 circulating water bath equipped with a Haake PG 20 temperature programmer. Temperature was monitored using a Micro-therm 1006 thermometer and an S/N117.c temperature probe in physical contact with the protein solution.

All samples were in 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 1 mM EDTA; if there was a cysteine in the protein, 1 mM DTT was added to the sample just before thermal melting and the sample was heated to 65°C for 5 min. All protein concentrations were determined by absorbance at 280 nm in 6 M guanidine hydrochloride using a Hewlett Packard 8425A spectrophotometer assuming the known absorbance for the one tryptophan and four tyrosines in the molecule (Cantor and Schimmel, 1980).

The ellipticity (θ_{222}) of the samples at 2°C, whether cross-linked or not, is 23 700 °cm^2/dmol which suggests that the proteins are ~62% helical (Woody and Tinoco, 1967; Gans *et al.*, 1991). The calculated θ_{222} for a 96 amino acid α-helical protein is −38 000 °cm^2/dmol. The leucine zipper, assumed to extend from the first d position (in this case it is an

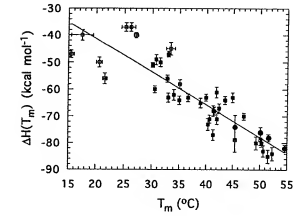

Fig. 14. The van't Hoff enthalpy per mole of dimer at the melting temperature [$\Delta H(T_m)$] is plotted versus the melting temperature (T_m) for four concentrations for EK_4 (●) and various mutant proteins (□). The error from the fitting procedure is shown for both parameters. There is a clear trend for those samples with higher T_m to show a larger $\Delta H(T_m)$. The assumption is that the main contribution to the temperature dependence of $\Delta H(T_m)$ for the various mutant proteins is Δc_p. The slope of the line through all the points was taken as the value of Δc_p (−1.16 ± 0.1 kcal/mol°) and used to calculate ΔG_0 at 37°C. Outliers from this trend exist; proteins containing a large number of unpaired basic residues, e.g. $R{\cdot}R_{34}$, $A{\cdot}R_{34}$, $K{\cdot}K_{34}$, show a smaller than expected $\Delta H(T_m)$, while those containing a large number of acidic residues, e.g. $E{\cdot}E_{34}$, $E{\cdot}A_{34}$, have a greater than expected $\Delta H(T_m)$.

isoleucine) all the way to the C-terminus, comprises 45% of the protein. The remaining 17% of unaccounted helicity is assumed to propagate into the basic region from the leucine zipper (Saudek *et al.*, 1991).

Calculation of thermodynamic parameters

Thermodynamic parameters were determined by curve fitting of the denaturation curves to the following equations using a non-linear least-squares fitting program (KaleidaGraph, Synergy Software). Ellipticity was normalized to fraction monomer using the equation:

$$\theta = (\theta_M - \theta_D)P_M + \theta_D \qquad (1)$$

where θ_M and θ_D represent the ellipticity values for the fully unfolded monomer and fully folded dimer species at each temperature. θ_M was found to be constant at the temperatures higher than the melting region for all the proteins studied. θ_D was approximated by a linear function of temperature $\theta_D = \theta_D(0) + \alpha T$. The fraction monomer ($P_M$) was expressed in terms of the equilibrium constant after solving the equation for a bimolecular reaction $2M \rightleftharpoons D$:

$$P_M = [(8KC + 1)^{1/2} - 1]/4KC \qquad (2)$$

where K is the equilibrium constant and C is the total protein concentration. K was assumed to be temperature dependent according to the equation

$$K = e^{-\Delta G/RT} \qquad (3)$$

This equation was expressed in terms of ΔH_m, the slope of the curve at T_m

$$K = C^{-1}\exp[\Delta H_m(1/T_m - 1/T)/R + \Delta c_p[T_m - T + T\ln(T/T_m)/R]] \qquad (4)$$

where T_m is the melting temperature ($P_M = 1/2$), ΔH_m is van't Hoff enthalpy of dimerization, Δc_p is the heat capacity change and R is the gas constant. Equations (1), (2) and (4) were combined and fitted to the CD data using KaleidaGraph fitting for the five parameters T_m, ΔH_m, α, θ_U and $\theta_D(0)$. We were unable to simultaneously solve for both ΔH_m and Δc_p because of the high degree of interdependence of these two variables. More satisfactory results were obtained by initially assuming Δc_p to be zero and then determining its value as described below. Δc_p could be calculated from the dependence of ΔH_m on T_m observed either by varying the concentration of a particular protein or using similar concentrations of different mutant proteins. Both methods gave similar results for Δc_p (Figure 14). This value of Δc_p has been used in our computations to calculate ΔG_0 at the reference temperature $T_o = 37°C$:

$$\Delta G_o = RT_o\ln C - \Delta H_m(T_o/T_m - 1) - \Delta c_p\,[T_m - T_o + T_o\ln(T_o/T_m)] \qquad (5)$$

The calculation of the thermodynamic parameters for heterodimers is complicated by the presence of three dimeric species. However, if an equimolar mixture of two proteins has a T_m which is significantly higher than that of either homodimer, it is possible to determine thermodynamic parameters of the heterodimer from the melting curve. In this case, the protein concentration (C) in equation (4) is the concentration of each protein. As a result, the total protein concentration in the mixing experiments was twice that in homodimer experiments.

The choice of the fitting interval from the CD melting curves affects the obtained thermodynamic parameters and results in an error of ΔG_0 ranging from 0.1 to 0.4 kcal/mol. Three independent protein preparations of the same sample ($E{\cdot}K_{1234}C$-S) have a ΔG_0 range of 0.17 kcal/mol.

Calculation of heterodimer formation

When two proteins, capable of homo- and heterodimerization, are present in solution, the following three reactions take place:

$$2M_1 \rightleftharpoons D_1, \quad [D_1]/[M_1]^2 = k_1 = e^{-\Delta G_1/RT} \qquad (6)$$

$$2M_2 \rightleftharpoons D_2, \quad [D_2]/[M_2]^2 = k_2 = e^{-\Delta G_2/RT} \qquad (7)$$

$$M_1 + M_2 \rightleftharpoons HD, \quad [HD]/[M_1][M_2] = k_3 = 2e^{-\Delta G^o HD/RT} \qquad (8)$$

(Note that in the last equation, ΔG^o_{HD} stands for an 'intrinsic' free energy of heterodimer formation. In this case, the exponential term is multiplied by two because two of four monomer collisions favor heterodimer while only one gives rise to each homodimer.) Because reactions (1)–(3) are linked, the concentration of heterodimer is related to the concentrations of homodimers:

$$[HD] = 2([D_1][D_2])^{1/2}\,e^{-\Delta G_{specificity}/RT} \qquad (9)$$

where $\Delta G_{specificity} = \Delta G^o_{HD} - (\Delta G_1 + \Delta G_2)/2$.

Equation (4) shows that for the preferential formation of heterodimer, $\Delta G_{specificity}$ should be large and negative. If homodimers and heterodimer are equally stable, $\Delta G_{specificity} = 0$, thus $[HD] = 2([D_1][D_2])^{1/2}$. Therefore, the heterodimer concentration can never exceed the sum of concentrations of homodimers, in the case of equal homodimer concentrations we will have the well-known ratio describing indifferent association with the heterodimer being twice as abundant as either homodimer $[D_1]:[HD]:[D_2] = 1:2:1$. The same relationship between heterodimer formation and dimer concentration will be exhibited by molecules which dimerize with different stability, but contribute to the stability of heterodimer independently (the heterodimer has no novel interactions compared to the homodimers) so that $\Delta G^o_{HD} = (\Delta G_1 + \Delta G_2)/2$ and $\Delta G_{specificity} = 0$ (Figure 11). When the heterodimer contains novel interactions when comparing to homodimers, its concentration could be either much greater (if $\Delta G_{specificity}/RT << 0$, see Figure 12) or much less (if $\Delta G_{specificity}/RT >> 0$) than the concentrations of homodimers. Knowing the dimerization free energy for given proteins, we will be able to predict the ratio $[HD]/2([D_1]*[D_2])^{1/2}$, reflecting the specificity of dimerization.

Acknowledgements

We thank Ernesto Friere, Kelly Thompson and the Biocalorimetry Center for use of the spectropolarimeter (NIH RR-04328), Mark Lewis for help with the XLA, and Mike Fritch and Carl Wu for the mass spectrometer analysis. We thank B.K.Lee and B.Garcia-Moreno for thoughtful discussions throughout this work, and V.Vinson, A.S.H.Goldman and members of the laboratory for comments on the manuscript. We thank J.Eldridge for the oligonucleotide synthesis and Claude Klee for encouragement and support.

References

Agre,P., Johnson,P. and McKnight,S. (1989) *Science*, **246**, 922–925.
Alber,T. (1992) *Curr. Opin. Genet. Dev.*, **2**, 205–210.
Alberti,S., Oehler,S., Wilcken-Bergmann,B.V. and Muller-Hill,B. (1993) *EMBO J.*, **12**, 3227–3236.
Amati,B., Brooks,M., Levy,N., Littlewood,T., Evan,G. and Land,H. (1993) *Cell*, **72**, 233–246.
Baxevanis,A. and Vinson,C. (1993) *Curr. Opin. Genet. Dev.*, **3**, 278–285.
Blaber,M., Zhang,X.-J. and Matthews,B. (1993) *Science*, **260**, 1637–1640.
Cantor,C. and Schimmel,P. (1980) *Biophysical Chemistry*. W.H.Freeman, New York.
Cao,Z., Umek,R. and McKnight,S. (1991) *Genes Dev.*, **5**, 1538–1552.
Carter,P., Winter,G., Wilkinson,A. and Fersht,A. (1984) *Cell*, **38**, 835–840.
Cohen,C. and Parry,D. (1990) *Protein*, **7**, 1–14.
Cooper,T. and Woody,R. (1990) *Biopolymers*, **30**, 657–676.
Crick,F. (1953) *Acta. Crystallogr.*, **6**, 689–697.

Dao-pin,S., Sauer,U., Nicholson,H. and Matthews,B. (1991) *Biochemistry*, **30**, 7142–7153.
Ellenberger,T., Brandl,C., Struhl,K. and Harrison,S. (1992) *Cell*, **71**, 1223–1237.
Fairman,R., Shoemaker,K., York,E., Stewart,J. and Baldwin,R. (1990) *Biophys. Chem.*, **37**, 107–119.
Gans,P., Lyu,P., Manning,M., Woody,R. and Kallenbach,N. (1991) *Biopolymers*, **31**, 1605–1614.
Gentz,R., III, Rauscher,J.F., Abate,C. and Curran,T. (1989) *Science*, **243**, 1695–1699.
Graddis,T., Myszka,D. and Chaiken,I. (1993) *Biochemistry*, 32, 12664–12671.
Hai,T. and Curran,T. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3720–3724.
Hai,T., Liu,F., Coukos,W. and Green,M. (1989) *Genes Dev.*, **3**, 2083–2090.
Harbury,P., Zhang,T., Kim,P. and Abler,T. (1993) *Science*, **262**, 1401–1407.
Ho,S., Hunt,H., Horton,R., Pullen,J. and Pease,L. (1989) *Gene*, **77**, 51–59.
Hodges,R., Sodak,J., Smillie,L. and Jurasek,L. (1972) *Cold Spring Harbor Symp. Quant. Biol.*, **37**, 299–310.
Horovitz,A., Serrano,L., Avron,B., Bycroft,M. and Fersht,A. (1990) *J. Mol. Biol.*, **216**, 1031–1044.
Hu,J., O'Shea,E., Kim,P. and Sauer,R. (1990) *Science*, **250**, 1400–1403.
Hu,J., Newell,N., Tidor,B. and Sauer,R. (1993) *Protein Sci.*, **2**, 1072–1084.
Ivashkiv,L.B., Liou,H.-C., Kara,C.J., Lamph,W.W., Verma,I.M. and Glimcher,L.H. (1990) *Mol. Cell. Biol.*, **10**, 1609–1621.
Iyer,S., Davis,D., Seal,S. and Burch,J. (1991) *Mol. Cell. Biol.*, **11**, 4863–4875.
Konig,P. and Richmond,T. (1993) *J. Mol. Biol.*, **233**, 139–154.
Kouzarides,T. and Ziff,E. (1988) *Nature*, **336**, 646–651.
Kouzarides,T. and Ziff,E. (1989) *Nature*, **340**, 568–571.
Laemmli,U. (1970) *Nature*, **277**, 680–685.
Landschultz,W., Johnson,P. and McKnight,S. (1988) *Science*, **240**, 1759–1764.
Landschultz,W., Johnson,P. and McKnight,S. (1989) *Science*, **243**, 1681–1688.
Loriaux,M., Rehfuss,R., Brennan,R. and Goodman,R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 9046–9050.
Lovejoy,B., Choe,S., Cascio,D., McRorie,D., DeGrado,W. and Eisenberg,D. (1993) *Science*, **259**, 1228–1293.
Lye,P., Liff,M., Marky,L. and Kallenbach,N. (1990) *Science*, **250**, 669–673.
Marqusee,S. and Baldwin,R. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8898–8902.
Merutka,G. and Stellwagen,E. (1991) *Biochemistry*, **30**, 1591–1594.
Mueller,C., Maire,P. and Schibler,U. (1990) *Cell*, **61**, 279–291.
Murre,C., McCaw,P. and Baltimore,D. (1989) *Cell*, **56**, 777–783.
Nicklin,M. and Casari,G. (1991) *Oncogene*, **6**, 173–179.
Nilges,M. and Brunger,A. (1991) *Protein Eng.*, **4**, 649–659.
O'Neil,K. and DeGrado,W. (1990) *Science*, **250**, 646–651.
O'Shea,E., Rutkowski,R., Stafford,W. and Kim,P. (1989) *Science*, **254**, 539–544.
O'Shea,E., Klemm,J., Kim,P. and Abler,T. (1991) *Science*, **254**, 539–544.
O'Shea,E., Rutkowski,R. and Kim,P. (1992) *Cell*, **68**, 699–708.
O'Shea,E., Lumb,K. and Kim,P. (1993) *Curr. Biol.*, **3**, 658–667.
Padmanabhan,S., Marqusee,S., Ridgeway,T., Laue,T. and Baldwin,R. (1990) *Nature*, **344**, 268.
Pu,W. and Struhl,K. (1993) *Nucleic Acids Research*, **21**, 4348–4355.
Robinson,C. and Sligar,S. (1993) *Protein Sci.*, **2**, 826–837.
Roman,C., Platero,J., Shuman,J. and Calame,K. (1990) *Genes Dev.*, **4**, 1404–1415.
Sanger,F., Nicklen,S. and Coulson,A. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Saudek,V., Pastore,A., Castiglione-Morelli,M., Frank,R., Gausepohl,H. and Gibson,T. (1991) *Protein Eng.*, **4**, 519.
Schindler,U., Menkens,A.E., Beckman,H., Ecker,J.R. and Cashmore,A.R. (1992) *EMBO J.*, **11**, 1261–1273.
Schmidt-Dor,T., Oertel-Buchheit,P., Pernelle,C., Bracco,L., Schnarr,M. and Granger-Scharr,M. (1991) *Biochemistry*, **30**, 9657–9664.
Scholtz,J., Qian,H., Robbins,V. and Baldwin,R. (1993) *Biochemistry*, **32**, 9668–9676.
Schuermann,M., Hunter,J., Hennig,G. and Muller,R. (1991) *Nucleic Acids Res.*, **19**, 739–746.
Serrano,L., Horovitz,A., Avron,B., Bycroft,M. and Fersht,A. (1990) *Biochemistry*, **29**, 9343–9352.
Studier,F. and Moffatt,B. (1986) *J. Mol. Biol.*, **189**, 113–130.
Thompson,K., Vinson,C. and Freire,E. (1993) *Biochemistry*, **32**, 5491–5496.
Turner,R. and Tjian,R. (1989) *Science*, **243**, 1689–1694.
Vinson,C., Sigler,P. and McKnight,S. (1989) *Science*, **246**, 911–916.
Vinson,C., Hai,T. and Boyd,S. (1993) *Genes Dev.*, **7**, 1047–1058.
Williams,S., Cantwell,C. and Johnson,P. (1991) *Genes Dev.*, **5**, 1553–1567.
Woody,R. and Tinoco,I. (1967) *J. Chem. Phys.*, **46**, 4927–4945.
Zamyatnin,A. (1984) *Annu. Rev. Biophys. Bioeng.*, **13**, 145–165.
Zhou,N., Kay,C. and Hodges,R. (1992) *Biochemistry*, **31**, 5739–5746.
Zhou,N., Kay,C. and Hodges,R. (1993) *Biochemistry*, **32**, 3178–3187.

Received on February 11, 1994; revised on April 11, 1994